

09/327, 984

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:50:09 ; Search time 120.01 Seconds
(without alignments)
357.991 Million cell updates/sec

Title: US-09-724-296-38
Perfect score: 3055
Sequence: 1 MTTGLESLSLGRGAPTV.....HKLKYNHDLWENHEKLSLS:580

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3055	100.0	580	AAW74473	Human MED1 endonuc
2	2958.5	96.8	565	21 AAY44504	Human Delta228-UV
3	2194.5	71.8	439	20 AAY76548	Human ovarian tumo
4	1230	40.3	257	20 AAW8701	Secreted protein e
5	1194	6.4	219	21 AAG02051	Human secreted pro
6	148.5	4.9	326	21 AAG16859	Arabidopsis thalia
7	148.5	4.9	327	21 AAG16858	Arabidopsis thalia
8	148.5	4.9	352	21 AAG16857	Arabidopsis thalia
9	145	4.7	1644	18 AAW25049	BRCA2 cancer susce
10	145	4.7	3418	18 AAW19211	Human breast cance
11	145	4.7	3418	18 AAW23287	Human breast and o

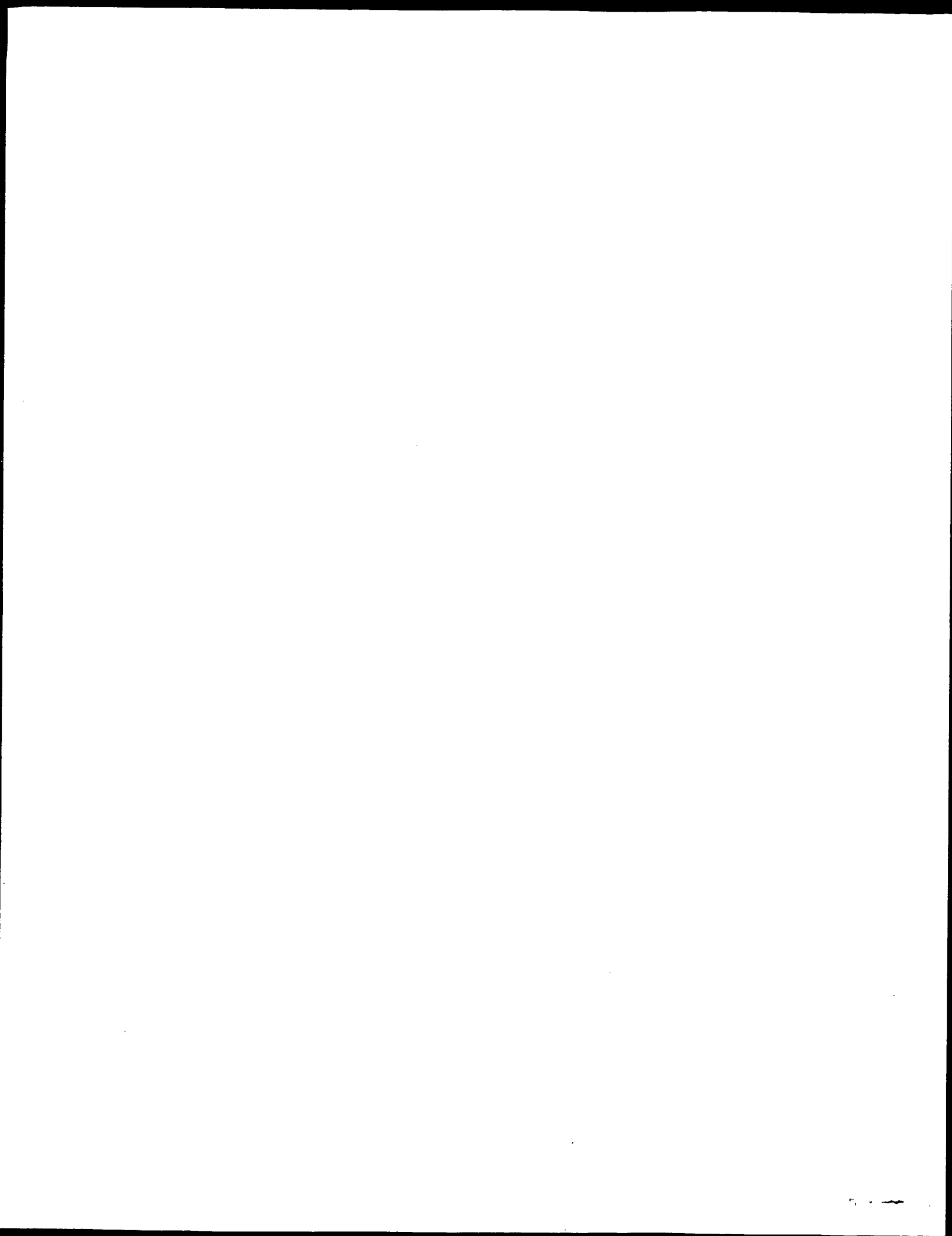
12	145	4.7	3418	20 AAY04356	Human BRCA2 (Omi3)
13	145	4.7	3418	20 AAY04357	Human BRCA2 (Omi4)
14	145	4.7	3418	20 AAY04358	Human BRCA2 (Omi5)
15	145	4.7	3418	20 AAY04354	Human BRCA2 (Omi1)
16	145	4.7	3418	21 AAY77819	BRCA2 protein sequ
17	144	4.7	3329	18 AAW25038	Partial BRCA2 can
18	143	4.7	3418	20 AAY04355	Human BRCA2 (Omi2)
19	141.5	4.6	782	11 AAB06991	Polypeptide antige
20	137	4.5	560	21 AAB19182	Lipid metabolism p
21	131.5	4.3	285	20 AAY14200	Mouse DNA demethyl
22	131.5	4.3	291	20 AAY14198	Human DNA demethyl
23	131	4.3	754	19 AAW33811	Tat stimulatory fa
24	130.5	4.3	411	20 AAY14197	Human DNA demethyl
25	130.5	4.3	411	22 AAB99915	Human protein sequ
26	130.5	4.3	411	22 AAG64314	Heart muscle cell
27	130.5	4.3	411	22 AAG64844	Expressed antigen
28	130	4.3	649	20 AAW90005	H. pylori GHPO 175
29	130	4.3	867	19 AAW98564	H. pylori GHPO 175
30	130	4.3	1743	19 AAW98879	Human prostate tum
31	127.5	4.2	281	20 AAY73829	Human prostate can
32	127.5	4.2	281	20 AAY48439	Human secreted pro
33	126	4.1	263	19 AAW74980	Mouse DNA demethyl
34	123	4.0	414	20 AAY14199	Peptide #4905 enco
35	122.5	4.0	995	22 AAM06223	Peptide #4097 enco
36	122.5	4.0	1202	22 AAM17663	Arabidopsis thalia
37	121.5	4.0	1024	21 AAG42350	Arabidopsis thalia
38	121.5	4.0	1105	21 AAG42349	Arabidopsis thalia
39	121.5	4.0	1189	21 AAG42348	Canine factor VIII
40	120.5	3.9	2343	20 AAW80989	Peptide #967 enco
41	120.5	3.9	2665	22 AAM14533	Peptide #987 enco
42	120.5	3.9	2665	22 AAM26950	Peptide #941 enco
43	120.5	3.9	2665	22 AAM02259	Human OREF ORF2255
44	120.5	3.9	3266	21 AAB42491	Plasmodium falcipa
45	119	3.9	646	21 AAB18188	

ALIGNMENTS

RESULT 1
AAW74473
ID AAW74473 standard; Protein; 580 AA.
XX AAW74473;
AC AAW74473;
XX
DT 19-MAY-1999 (first entry)
XX
XX Human MED1 endonuclease protein sequence.
XX
DE
XX
XX
XX
KW Endonuclease; MED1; human; methyl-CpG binding endonuclease-1;
KW DNA fidelity; DNA manipulation; cancer; fragile X syndrome; therapy;
KW myotonic dystrophy; Huntington's disease; spinocerebellar ataxia;
KW Kennedy's disease; triplet repeat expansion disorder.
XX
XX Homo sapiens.
XX
XX WQ9904636-AL.
XX
XX 04-FEB-1999.
XX
PF 28-JUL-1998; 98WO-US15828.
XX

PR 28-JUL-1997; 97US-0053936.
XX
PA (FOX-) FOX CHASE CANCER CENT.
XX
XX
PI Bellacosa A;
XX
XX WPI; 1999-142462/12.
DR N-PSDB; AAX22002#.
XX
PT New nucleic acid encoding human endonuclease MED1 involved in DNA mismatch repair - used for diagnosing susceptibility to cancer and

Aug Nov 28/2000



fragile X syndrome, and therapeutically

Claim 8; Fig 3; 109pp; English.

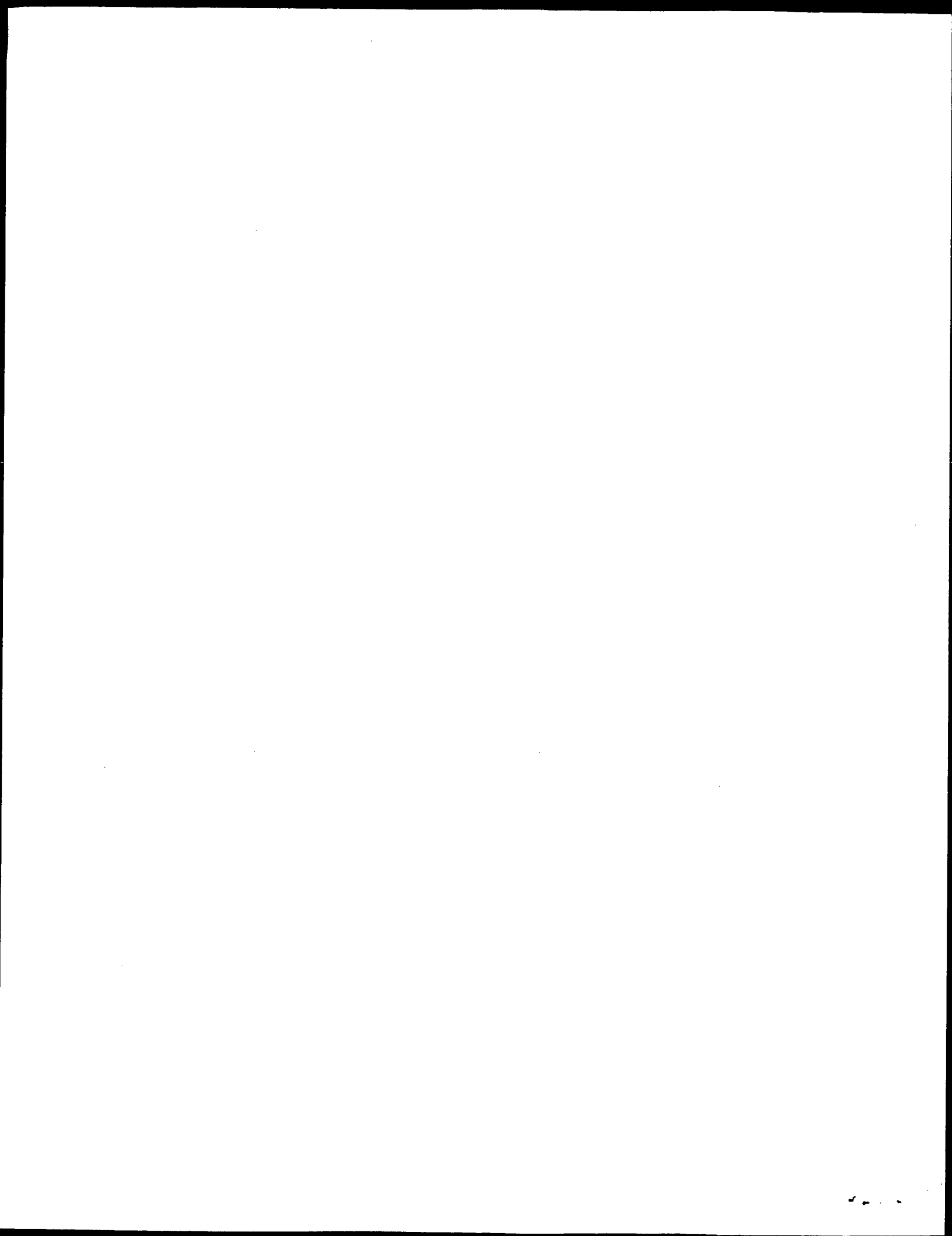
This sequence is the human MED1 endonuclease of the invention. MED1 (for methyl-CpG binding endonuclease-1) is used to screen for specific modulators (potential therapeutic agents particularly mimetics of MED1) and to study interactions involved in maintaining DNA fidelity, for DNA manipulation and to raise antibodies. Susceptibility or predisposition to cancer (particularly colorectal or endometrial, especially hereditary non-polyposis colorectal cancer), or its prognosis, where caused by alterations in the MED1-encoding gene, are identified by sequence comparison, amplification, detecting altered polypeptide, and restriction fragment mapping, hybridisation (particularly to probes specific for a mutant allele). These same methods can also be used to diagnose fragile X syndrome and other diseases (e.g. myotonic dystrophy, Huntington's disease, spinocerebellar ataxia and Kennedy's disease) associated with triplet repeat expansion. The DNA, or its fragments, are used as probes and primers in the above diagnostic methods, also to isolate homologous sequences, as sources of antisense sequences and for gene transfer, particularly to restore drug sensitivity to drug-resistant cancer cells.

Sequence 580 AA:

Query Match 100.0%; Score 3055; DB 20; Length 580;
Best Local Similarity 100.0%; Pred. No. 7.6e-284;
Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGTGLESLSLGRGAAPTTSERLVPDPNDLRKEDVAMELERVGEDEEQMKRSSE 60
DB 1 mgttgleslsldrgaaptvtasserlvppndlrkedvamelervgedeeqmmikrsse 60
QY 61 CNPLQEPPIASAGFAGTAGTECKRSVPCGWVERVKQRLFGKTAGRFDVYFISPOGLKFRS 120
DB 61 cnpllqepiasagfagtagteckrsvpcgwerlvkqrlfgktagrfdvyfispqglkfrs 120
QY 121 KSSLANYLHNKGETSLKPEDFDTVLKRGKISRYKDCSMAALTSHLQNSNNWNLRT 180
DB 121 ksslanylhnkgetslkpedfdftvlksrgikrsykdcsmaaltshlqnsnnwnlrt 180
QY 181 RSKCKDVFMPSPSSSELSQESRGLSNFTSHLLKDEGVDVDFNFRKVRKPKGVITLKG 240
DB 181 rskckdvfmppsssselsqesrglsnftshllkdegvddvfnfrkvrkpkgvitlkg 240
QY 241 IPKTKKGRKSCSGFVQSDSKRESVCKADAEPVAKQSOLDRTVCISDAGACGETL 300
DB 241 ipkttkgrkscsgfvqsdskresvckadaesepvacksgldrtvcisdagacgetl 300
QY 301 SVTSEENSLVKKERSLSSGNFCSQKTSGLINKFCSAKDSEHNEKYEDTLESEETG 360
DB 301 svtseenslvkkerslssgnfcsqktsvglinkfcsakdsehnekyedtleseetg 360
QY 361 KVEVERKEHLHTDILKRGSEMDNNSCPTKDFTEGKIFQEDTIPRTQIERKTSLYFSS 420
DB 361 kveverkehldilkrsemdnncsptrkdftegekifqedtiprtqierktslyfss 420
QY 421 KYNKEALSPRRKAFKWTTPSPFNLVQETLPHDPWKLLIATIFLNRTSKGMAIPVLWK 480
DB 421 kynkealspprrkafkwtppspfnlvqetlphdpwklliatiflnrtskgmaipvlwk 480
QY 481 FLEKYPYSAEVRTADNRDYSLELKLPLGLYDLRAKTIVKFSDEVLTWKOWKYPTELHGIGY 540
DB 481 flekypysaevrtadnrdsylelklplgllydlraktivkfsdevltwkowkyptelhgigy 540
QY 541 GNDYSYIFCVNEKQVHPEDHLKYNKYHDWLNWENHEKLSLS 580
DB 541 gndysyifcvnekwqvhpedhlkynkyhdwlnwenheklsls 580

ID RAY44504 standard; Protein; 565 AA.
XX AC RAY44504;
XX DT 27-MAR-2000 (first entry)
XX DE Human delta228-UV damage endonuclease.
XX KW Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide; glutathione-S-transferase signal peptide; uvell+ gene product; UV irradiation; DNA damage; UV radiation damage; photoproduct; KW abasic site; platinum diadduct; mismatched nucleotide pairing; KW nucleotide alkylation; skin cancer.
XX OS Homo sapiens.
XX PN WO9963828-A1.
XX PD 16-DEC-1999.
XX PF 08-JUN-1999; 99WO-US12910.
XX PR 08-JUN-1998; 98US-0088521.
PR 18-MAY-1999; 99US-0134752.
XX PA (UYEM-) UNIV EMORY.
XX PI Doetsch PW, Kaur B, Avery AM;
XX DR WPI; 2000-116417/10.
XX PT A new truncated ultraviolet damage endonuclease for treatment of skin cancers -
XX PS Claim 16; Page 60; 133pp; English.
XX CC The present sequence is human delta228-UV damage endonuclease.
CC Delta228-UVDE is a 288 amino acid deletion of the N-terminal of the uvell+ gene product. This is expressed in frame with a GST leader sequence to generate a fusion protein. This provides stable endonuclease fragments for cleaving a double-stranded DNA molecule that has a distorted structure resulting from UV radiation damage, a photoproduct, an abasic site, mismatched nucleotide pairing, platinum diadduct, an intercalated molecule or alkylation of a nucleotide. Uvell can be used in compositions for internal or topical application and as a therapeutic agent for skin cancers.
XX SQ Sequence 565 AA;
Query Match 96.8%; Score 2958.5; DB 21; Length 565;
Best Local Similarity 97.4%; Pred. No. 1.3e-274;
Matches 565; Conservative 0; Mismatches 0; Indels 15; Gaps 1;
QY 1 MGTGLESLSLGRGAAPTTSERLVPDPNDLRKEDVAMELERVGEDEEQMKRSSE 60
DB 1 mgttgleslsldrgaaptvtasserlvppndlrkedvamelervgedeeqmmikrsse 60
QY 61 CNPLQEPPIASAGFAGTAGTECKRSVPCGWVERVKQRLFGKTAGRFDVYFISPOGLKFRS 120
DB 61 cnpllqepiasagfagtagteckrsvpcgwerlvkqrlfgktagrfdvyfispqglkfrs 120
QY 121 KSSLANYLHNKGETSLKPEDFDTVLKRGKISRYKDCSMAALTSHLQNSNNWNLRT 180
DB 121 ksslanylhnkgetslkpedfdftvlksrgikrsykdcsmaaltshlqnsnnwnlrt 180
QY 181 RSKCKDVFMPSPSSSELSQESRGLSNFTSHLLKDEGVDVDFNFRKVRKPKGVITLKG 240
DB 181 rskckdvfmppsssselsqesrglsnftshllkdegvddvfnfrkvrkpkgvitlkg 240
QY 241 IPKTKKGRKSCSGFVQSDSKRESVCKADAEPVAKQSOLDRTVCISDAGACGETL 300
DB 241 ipkttkgrkscsgfvqsdskresvckadaesepvacksgldrtvcisdagacgetl 300




```

Best Local Similarity 100.0%; Pred. No. 7.9e-190;
Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRLKNIQISKRIVFTILKQAFKGNHPCVSVCTITYSRFHCLPDLTKSLPMSKT 60
DB 1 MLRLKNIQISKRIVFTILKQAFKGNHPCVSVCTITYSRFHCLPDLTKSLPMSKT 60
QY 61 TSLMLPQVNTGANSFSAETPVDLKENETELANISGPHKSTSTSTRKRARSSKKKATDS 120
DB 61 TSLMLPQVNTGANSFSAETPVDLKENETELANISGPHKSTSTSTRKRARSSKKKATDS 120
QY 121 VSDKIDSVASYDSTHLRSSRSKKPVNNSSESESESEQISKATKYKQKEEEVEE 180
DB 121 VSDKIDSVASYDSTHLRSSRSKKPVNNSSESESESEQISKATKYKQKEEEVEE 180
QY 181 VDESKLNESSDDEFVPEQLETPISKRRRSKAKNEKESTMNLDHAPREMFDC 240
DB 181 VDESKLNESSDDEFVPEQLETPISKRRRSKAKNEKESTMNLDHAPREMFDC 240
QY 241 DKPTIPWRGRGLYACNLTLRSKKEVFCSTRTITTTORGLSVKQLGTONVLDLKL 300
DB 241 DKPTIPWRGRGLYACNLTLRSKKEVFCSTRTITTTORGLSVKQLGTONVLDLKL 300
QY 301 EWNHNFGLHFRVSSDLFPFASHAKYGYTLLEFAOSHLEEVGKLANKYNHRLTWHPGGYTQ 360
DB 301 EWNHNFGLHFRVSSDLFPFASHAKYGYTLLEFAOSHLEEVGKLANKYNHRLTWHPGGYTQ 360
QY 361 IASPREVVVDSAIRDLAYHDEILSRMKLNEQLNKDAVLIHLGTFEGKKTLDLRFK 420
DB 361 IASPREVVVDSAIRDLAYHDEILSRMKLNEQLNKDAVLIHLGTFEGKKTLDLRFK 420
QY 421 QRLSDSVKARLVLENDVSWSVQDILLPLCQELNPLVLDWHHNNIVPGLTREGSLDML 480
DB 421 QRLSDSVKARLVLENDVSWSVQDILLPLCQELNPLVLDWHHNNIVPGLTREGSLDML 480
QY 481 IPTIRETWRTKIGITQKHYSADPTAISGMKRAHSDRVDFPDCPTMDLMEAEKE 540
DB 481 IPTIRETWRTKIGITQKHYSADPTAISGMKRAHSDRVDFPDCPTMDLMEAEKE 540
QY 541 QAVFELCRRYELQNPCCPLETIMGPEYDTRDGYPPGAEKRLTARKRRSRKEEVEDEK 599
DB 541 QAVFELCRRYELQNPCCPLETIMGPEYDTRDGYPPGAEKRLTARKRRSRKEEVEDEK 599

RESULT 2
Q01408 PRELIMINARY; PRT; 656 AA.
AC Q01408;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-MAY-1999 (TremBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TremBLrel. 10, Last annotation update)
DE UV-ENDONUCLEASE.
GN UVEL.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A;
RX MEDLINE=95292980; PubMed=7774597;
RA Yajima H., Takao M., Yasuhira S., Zhao J.H., Ishii C., Inoue H.,
RA Yasui A.;
RT "A eukaryotic gene encoding an endonuclease that specifically repairs
RT DNA damaged by ultraviolet light.";
RL EMBO J. 14:2393-2399(1995).
DR EMBL; D11392; CAB20113.1;
KW Endonuclease.
SQ SEQUENCE 656 AA; 74413 MW; 4B231E7A8735C86F CRC64;

Matches 241; Conservative 88; Mismatches 149; Indels 181; Gaps 18;
QY 110 ARSSKKKATDSVSKIDSVASYDSST-----HLRRSR-----SKPVNNSSESE- 157
DB 3 SRKSKAAALDTPQSESTFSSTLDSSAPARNLRSGRNILQPSSEKDRDHEKRSGEEL 62
QY 158 -----SEEQISKATKKVKQKEEEVEEVEDEK 184
DB 63 AGRMMGKQANGHCLREGKEQEGVKMAIEGLARMERLRQATKQKKOLEED----- 114
QY 185 SLKNSSDDEFVPEQLETPISKRRRS-----RSSAKNEKESTMNLDHHA----- 232
DB 115 GIPVSVSRF-PTAPYHHKSTINAEERAEKPEVLKTHSKDVEREAEIGDQVWMEPAAT 173
QY 233 ----PREMFDCLDK-----PIWRGRGLYACNLTLRSKKEVFCSTRTCIT 275
DB 174 NIEPEDAQAADAEARGAPPAVNSSYLPLPKRGLYACNLTYLRNAKAPPISFSSRTCA 233
QY 276 TI-----ORDGLSVKQLGTONVLDLILKLVENHNFG 307
DB 234 SIVDHRHPLOFEDEPEHHLKKNPKDSKEPQDELGHKFVQELGLANARDIVKMLCWNKYG 293
QY 308 IHFMRVSSDLFPFASHAKYGYTL-EFAOSHLEEVGKLANKYNHRLTWHPGGYTQIASPRE 366
DB 294 IREFLRSSEMFPPASHPVGHYKGLAPFASEVLAEGRAVAELGHRLTTHPGQFTQLGSPRK 353
QY 367 VVDSAIRDLAYHDEILSRMKLNEQLNKDAVLIHLGTFEGKKTLDLRFKKNYORLSDS 426
DB 354 EVESAIRDLAYHDEILSRMKLNEQLNKDAVLIHLGTFEGKKTLDLRFKKNYARLSQS 413
QY 427 VKARLVLENDVSWSVQDILLPLCQELNPLVLDWHHNNIV--PGLTREGSLDLM--PLIP 482
DB 414 CKNLVLENDVGVTVHDLPLVCEELNIPMLVDYHHNNICFDPALHREGTLDISDKLQE 473
QY 483 TIRETWRTKIGITQKHYSADPTAISGMKRAHSDRVDFPDCPTMDLMEAEKEQEA 542
DB 474 RIANTWKRGITQKHYSADPTAISGMKRAHSDRVDFPDCPTMDLMEAEKEQEA 532
QY 543 VFELCRRYEL-----QNPCCP-----LEIMGPE--- 565
DB 533 VFELMTFRLPOFEKINDMVPYDRDDENPAPVPKAPKKKGGKRRKRTTDEAAPEEVD 592
QY 566 --YDQTRDG-----YPPGAEKRLTARKRRSRK-----EEVEED 597
DB 593 TAADDVKADEPGKPEVEBERAMGGPYNVWNPPLGCEEWLKPCKREVKKGVPEVEDE 651

RESULT 3
Q014F0 PRELIMINARY; PRT; 317 AA.
AC Q014F0;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TremBLrel. 16, Last annotation update)
DE PUTATIVE UV-ENDONUCLEASE.
GN YWJD.
OS Bacillus cereus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1396;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14579 TYPE STRAIN;
RX MEDLINE=20055637; PubMed=10589720;
RA Okstad O., Gominet M., Purnelle B., Rose M., Lereclus D., Kolsto A.B.;
RT "Sequence analysis of three Bacillus cereus loci under P1cR-regulated
RT genes encoding degradative enzymes and enterotoxin.";
RL Microbiology 145:3129-3138(1999).
DR EMBL; AJ243712; CAB69813.1;
KW Endonuclease.
SQ SEQUENCE 317 AA; 36990 MW; A6AC35F5800A22C3 CRC64;

```

Query Match 32.5%; Score 1011.5; DB 3; Length 656;
Best Local Similarity 36.6%; Pred. No. 2.2e-56;



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:52:23 ; Search time 74.94 Seconds
(without alignments)
666.807 Million cell updates/sec

Title: US-09-724-296-36
Perfect score: 3482
Sequence: 1 MPSRKSAAALDTPOSESST.....REVKKGKVPVEVEGEFGD 656

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

al number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

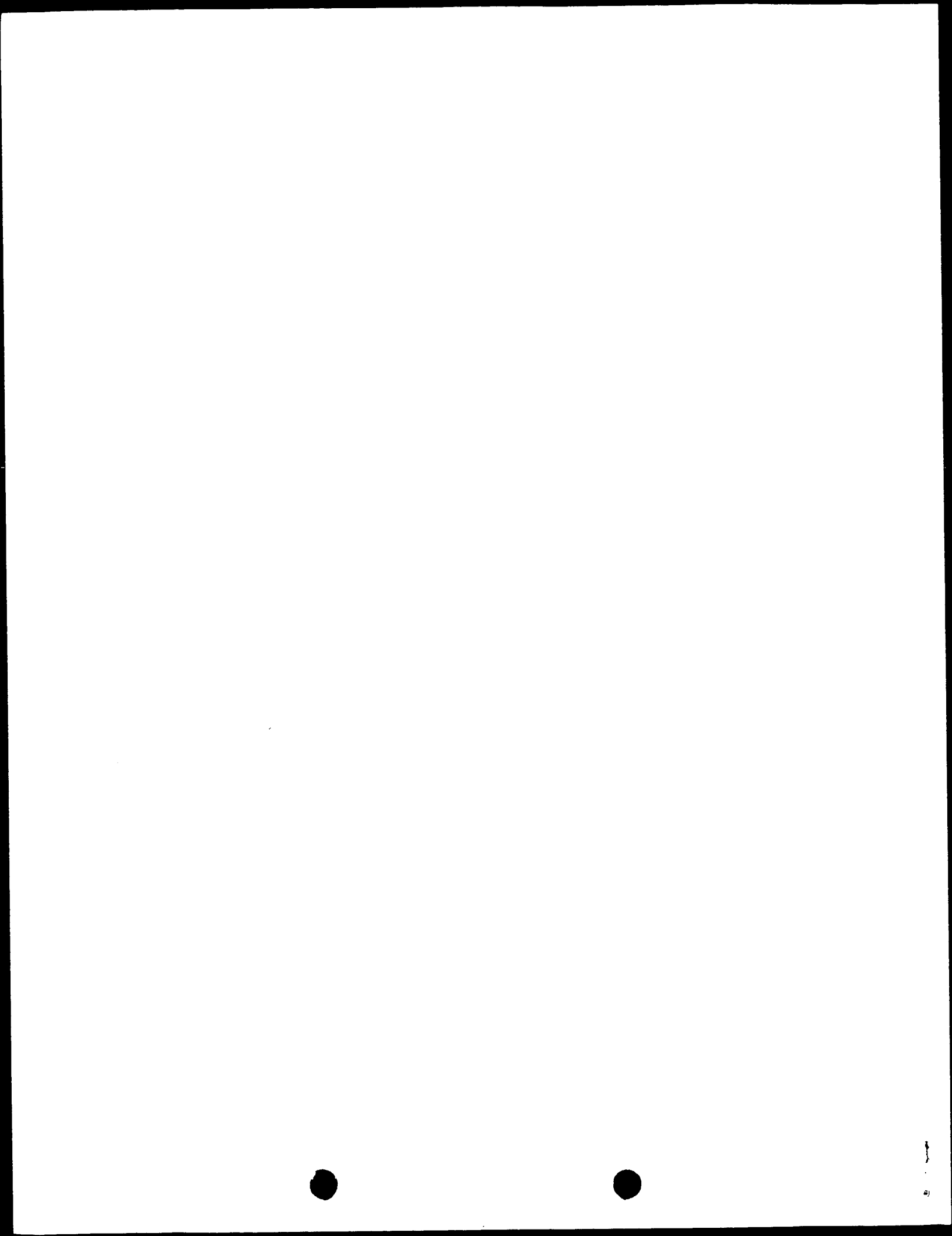
Result No.	Score	Query Match	Length	ID	Description
1	3472	99.7	656	S55262	UV-endonuclease -
2	1011.5	29.0	599	S71134	UV-endonuclease -
3	377	10.8	322	E83974	hypothetical prote
4	348.5	10.0	320	S55418	UV-endonuclease ho
5	261.5	7.5	326	C75350	probable UV damage
6	160.5	4.6	1280	T00365	hypothetical prote
7	155.5	4.5	5327	T13564	microtubule-associ
8	146	4.2	796	S56231	hypothetical prote
9	141.5	4.1	1110	S52116	NF-180 - sea lamp
10	140	4.0	727	T17292	hypothetical prote
11	139.5	4.0	1510	T33100	hypothetical prote
12	139	4.0	1815	C81169	IgA-specific metal
13	138	4.0	1547	T28657	blackjack protein, p
14	137	3.9	1560	T42727	cell proliferation
15	137	3.9	2897	B48666	cell proliferation
16	137	3.9	3256	A48666	neurofilament trip
17	136.5	3.9	845	A45669	neurofilament trip
18	136.5	3.9	849	S00030	neurofilament trip
19	136	3.9	3942	T42730	Bassoon protein -
20	135.5	3.9	971	T24866	hypothetical prote
21	135	3.9	793	JH0628	caldesmon - human
22	134	3.8	2663	S28261	centromere protein
23	133.5	3.8	980	E71606	hypothetical prote
24	132.5	3.8	915	S54485	CE51 protein - yea
25	132.5	3.8	6642	T29375	hypothetical prote
26	132	3.8	734	F85073	protein UNC-89 - C
27	131.5	3.8	1222	T22490	hypothetical prote
28	131	3.8	606	A43427	neurofilament trip
29	128.5	3.7	856	T16543	hypothetical prote

30 128.5 3.7 7962 2 I38346
31 128 3.7 1020 1 QFH0H
32 128 3.7 1805 2 A34736
33 128 3.7 3488 2 T34418
34 128 3.7 4151 2 T13734
35 127.5 3.7 2453 2 S60254
36 126.5 3.6 720 2 T36819
37 126.5 3.6 2416 2 T13825
38 126 3.6 522 2 C96608
39 126 3.6 1773 2 A81937
40 125.5 3.6 786 2 C86406
41 125.5 3.6 1400 2 T52359
42 125.5 3.6 2441 2 D71623
43 124.5 3.6 1829 2 T26135
44 124.5 3.6 4377 2 A55575
45 124 3.6 1072 1 A37221

ALIGNMENTS

RESULT 1
S55262
UV-endonuclease - Neurospora crassa
C:Species: Neurospora crassa
C:Date: 10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 20-Jun-2000
C:Accession: S55262
R:Yajima, H.; Takao, M.; Yasuhira, S.; Zhao, J.H.; Ishii, C.; Inoue, H.; Yasui, A.
EMBO J. 14, 2393-2399, 1995
A:Title: A eukaryotic gene encoding an endonuclease that specifically repairs DNA dam
A:Reference number: S55262; MUID:95292980
A:Accession: S55262
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-656 <YAU>
A:Cross-references: EMBL:D11392; NID:g1526560; PID:g927215
C:Genetics:
A:Gene: uvel

Query Match 99.7%; Score 3472; DB 2; Length 656;
Best Local Similarity 99.7%; Pred. No. 2.2e-208;
Matches 654; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Oy 1 MPSRKSAAALDTPOSESSTFSSDSSAPSPARNLRSGRNILQPSSEKDRDHEKRSGE 60
Db 1 MPSRKSAAALDTPOSESSTFSSDSSAPSPARNLRSGRNILQPSSEKDRDHEKRSGE 60
Oy 61 ELAGRMGKDANGHCLREGKEQEGVKMAIEGLARMERRLQRAKROKKOLEEDGIPVPS 120
Db 61 ELAGRMGKDANGHCLREGKEQEGVKMAIEGLARMERRLQRAKROKKOLEEDGIPVPS 120
Oy 121 VVSREPTAPYHHKSTNAEEREAKPEVLKTHSKOYERAEIGVDVVKMEPAATNIIEPED 180
Db 121 VVSREPTAPYHHKSTNAEEREAKPEVLKTHSKOYERAEIGVDVVKMEPAATNIIEPED 180
Oy 181 AODAEAGAAPPAVNSYPLPWKGRGLGACLTNTYLRNAKPPFSSRTCRMASIVDHRH 240
Db 181 AODAEAGAAPPAVNSYPLPWKGRGLGACLTNTYLRNAKPPFSSRTCRMASIVDHRH 240
Oy 241 PLOFEDEPEHHLKNDKSKPEQDELGHKFVQELGLANARDIVKMLCWNKYGIRFLRLS 300
Db 241 PLOFEDEPEHHLKNDKSKPEQDELGHKFVQELGLANARDIVKMLCWNKYGIRFLRLS 300
Oy 301 SEMPPFASHPVHGYKLPAPFASEVLAEEAGRAVAELGHLRLTHPGQFTQLGSPRKEVESAI 360
Db 301 SEMPPFASHPVHGYKLPAPFASEVLAEEAGRAVAELGHLRLTHPGQFTQLGSPRKEVESAI 360
Oy 361 RDLEYHDELLSLKLPQOONRDVMIHMGQFGDGAATLERFKRNYARLSQSKNRLVL 420
Db 361 RDLEYHDELLSLKLPQOONRDVMIHMGQFGDGAATLERFKRNYARLSQSKNRLVL 420
Oy 421 ENDDVGWTVHDLFPVCEELNIPMVLVDYHHNICFPAHLREGTLDISDPKLOERIANWK 480



Db 421 ENDVGTVDHLLPVCEELNIPVLDYHHNICFDPFAHLREGTLDISDPKQERIANTWK 480
 QY 481 RKGIKMHYSEPCDGAIVTPRRKRHRVMTLPCCPDMDLMIKAKKEQAVFELMRTF 540
 Db 481 RKGIKMHYSEPCDGAIVTPRRKRHRVMTLPCCPDMDLMIKAKKEQAVFELMRTF 540
 QY 541 KLPGFEKINDVYRDDNRRPAPVAPKPKKGGKRRKRTTDEAAPEEVDTAADVDK 600
 Db 541 KLPGFEKINDVYRDDNRRPAPVAPKPKKGGKRRKRTTDEAAPEEVDTAADVDK 600
 QY 601 APEGKPEVEERAMGGPNRYWPLGCEEWLKPCKREVKKGVPEEVEDEGEFDG 656
 Db 601 APEGKPEVEERAMGGPNRYWPLGCEEWLKPCKREVKKGVPEEVEDEGEFDG 656

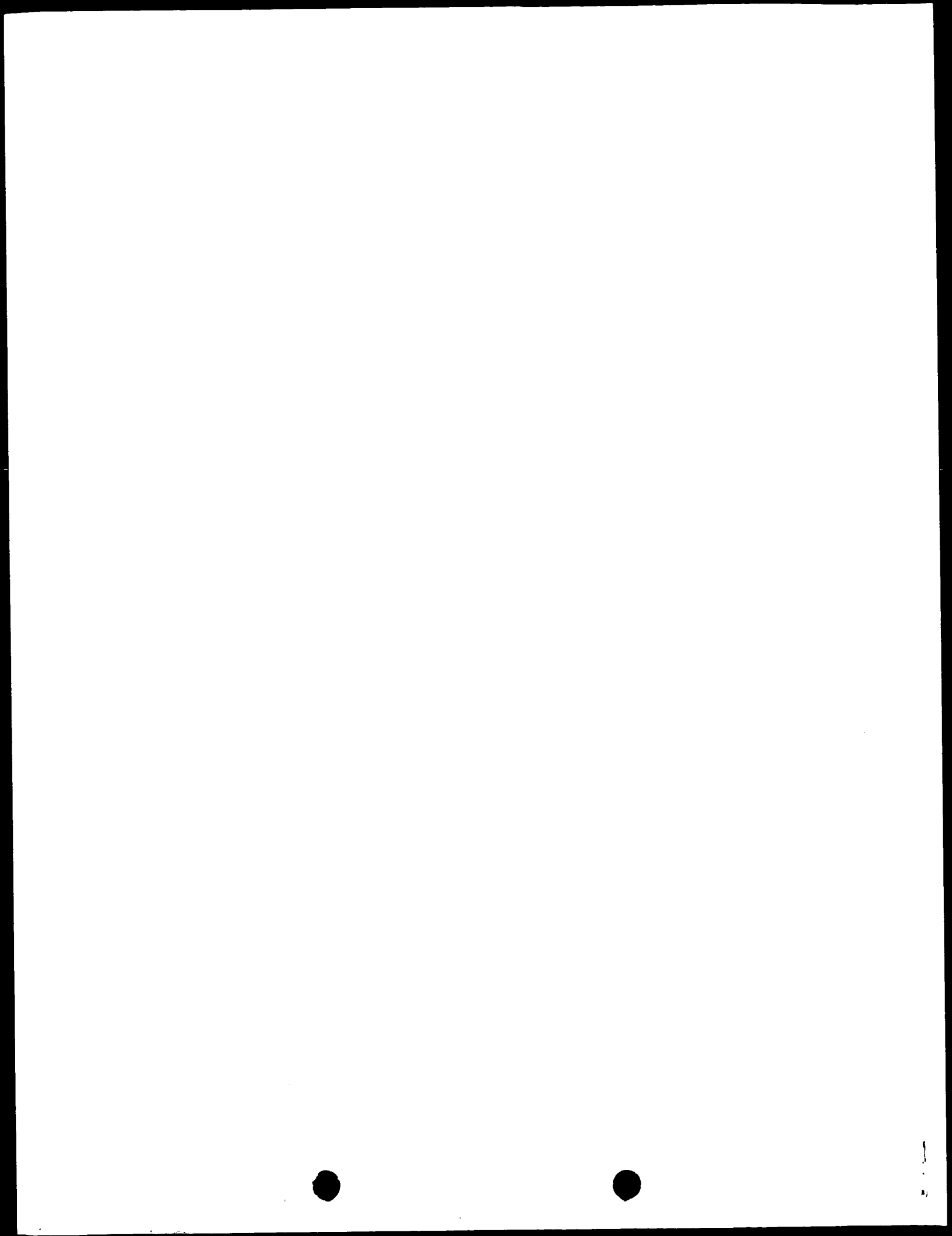
RESULT 2
 S71134
 UV-endonuclease - fission yeast (Schizosaccharomyces pombe)
 Species: Schizosaccharomyces pombe
 Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 20-Jun-2000
 Accession: S71134; T39815
 R.Takao, M.; Yonemasu, K.; Yamamoto, K.; Yasui, A.
 Nucleic Acids Res. 24, 1267-1271, 1996
 A:Title: Characterization of a UV endonuclease gene from the fission yeast Schizosaccharomyces pombe
 A:Reference number: S71134; MUID:96188860
 A:Accession: S71134
 A:Status: preliminary; nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-599 <TAK>
 A:Cross-references: EMBL:D78571; NID:g1399000; PIDN:BAAL1415.1; PID:g1399001
 A:Experimental source: strain SP972
 R.Lyne, M.; Rajandream, M.A.; Barrrell, B.G.; Lucas, M.; Gaillardin, C.
 submitted to the EMBL Data Library, June 1998
 A:Reference number: 221881
 A:Accession: T39815
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-599 <LYN>
 A:Cross-references: EMBL:AL023859; PIDN:CAAL19577.1; GSPDB:GN00067; SPDB:SPBC19C7.09c
 A:Experimental source: strain 972h; cosmid c19C7
 C:Genetics:
 A:Gene: UVPE; SPBC19C7.09c
 A:Map position: 2

Query Match 29.0%; Score 1011.5; DB 2; Length 599;
 Best Local Similarity 36.8%; Pred. No. 1.9e-55;
 Matches 241; Conservative 88; Mismatches 149; Indels 181; Gaps 18;
 QY 3 SRKSKAALDTPSESTSTLSSAPSPARNLRSGRNILQPSKEDRDHEKRSSEEL 62
 Db 110 ARSKKKATDSVDKIDESVASYDSST-----HLRSSR-----SKKPNYNSSESE- 157
 QY 63 AGRMWDGANGCHLRECKEGBEGVKNALGLARMERLRORATKROKLFEE----- 114
 Db 158 -----SEQISKATKKVKQKEEYVEEVDK 184
 QY 115 GIPVPSVYSRF-TPATYHHKSTNAEREAKEPVLTKTHSKDVEEAEIGVDVVVKMEPAT 173
 Db 185 SLKNSSSDEFPVVPQLETPISKRRS-----RSANLEKESTMNDLDDHA----- 232
 QY 174 NIEPEDAQDAERGAARPPAVNSYLPKPGKGLGYACILTYLRNAKPPIFSSRTCRMA 233
 Db 233 -----PREMFCDLCK-----PIPWGRGLGYACILTYLRNAKPPIFSSRTCRIT 275
 QY 234 SIYDHRHPLQFEDEPEHLKKNPKDSKEQDELGHKGFVQELGLANARQIVKMLCWNKYG 293
 Db 276 TI-----QRDGLSVRLQGTQNVLDLILKLVENHNFG 307
 QY 294 IRELRLSSEMPFPFASHVPHVHYKILAPPASEVLAEGRAVAAELGHLRTTHPGQFTQLGSPRK 353
 Db 308 IHPMYRSSDLFPFASHAKYGYTL-EFAQSHLEVGLKANKYNNHRLTMHPGQYTOIASPRE 366

QY 354 EVVESAIRDLEYHDELLSLKLPEQQNRDAMVIHMGQFGDKAATLERFKRNYARLSQS 413
 Db 367 VVDSAIRDLAYHDEILSRMLNEQLNKDAVLIHILGGTPEGKKTDLRFKKNYQRLSDS 426
 QY 414 CKNRLVLENDVGTVDHLLPVCEELNIPVLDYHHNICFDPFAHLREGTLDISDPKLOE 473
 Db 427 VKARLVLENDVSKSVQDILLPLCEELNIPVLDYHHNICFDPFAHLREGTLDISDPKLOE 473
 QY 474 RIANTWKRKGIKMHYSEPCD-GAVTPRRKRHRVMTLPCCPDMDLMIKAKKEQA 532
 Db 483 TIRETWRTKGIKMHYSEPCD-TAISGMKRAHSDRVDFPCDPTMDLMIKAKKEQA 542
 QY 533 VFELMRTFKLPGEKINDVYRDDNRRPAPVAPKPKKGGKRRKRTTDEAAPEEVD 592
 Db 543 VFELCRRYEL-----ONPPCP-----LEIMGPE----- 565
 QY 593 TAADDVKDAPGKPEVEERAMGGPNRYWPLGCEEWLKPCKREVKKGVPEEVEDE 651
 Db 566 --YDQTRDG-----YYPGGAEKRLTARKRSRK-----EVEED 597

RESULT 3
 E83974
 hypothetical protein BH2597 [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
 C:Accession: E83974
 R.Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Fujii, F.; H.
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans
 A:Reference number: A83650; MUID:20263314
 A:Accession: E83974
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-322 <STO>
 A:Cross-references: GB:AF001516; GB:BA000004; NID:g10175192; PIDN:BA06316.1; GSPDB:G
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH2597

Query Match 10.8%; Score 377; DB 2; Length 322;
 Best Local Similarity 29.1%; Pred. No. 2.4e-16;
 Matches 105; Conservative 61; Mismatches 127; Indels 68; Gaps 10;
 QY 207 RUGYACLTLYLRNAKPPIFSSRTCRMASIVDRHPLQFEDEPEHLKKNPKDSKEPQDEL 266
 Db 4 QGYVAMSWELANASP-----SKTMTAT-----QEKIEDH-----EA 36
 QY 267 GHKFVQELGLANARDIVKMLCWNKYGIRFLRLSSEMPFASHPV-HGYKLPAPFASEVLA 325
 Db 37 GLRKLERIAKTNLNCLRLKHLNLAYQISFFRLSSKLVPLVNHPLTEGWKYLELAIELO 96
 QY 326 EAGRVAAELGHLRTTHPGQFTQLGSPKRVVESAIRDLEYHDELLSLKLPEQQNRDAMV 385
 Db 97 AVGEFASHQMRIDFDPHFVYVLANSEAKITRRSLQTLHYKLLKGMIDPRHR----- 152
 QY 386 ITHMGQFGDKAATLERFKRNYARLSOSCKNRLVLENDVGTVDHLLPVCEELNIPWL 445
 Db 153 VLHVGKKGKGEAGLEQFIENTAPSKLSMLMLENDKSYTIDVLYLGERLAIPWL 212
 QY 446 DYHHNICFDPFAHLREGTLDISDPKLOERANTWKRKGIKMHYSEPCDGAIVTPR----- 501
 Db 213 DIHHHDVLHRSKSLQE-----TWQRIVATWEDSPLPVKIHLSPLSGEDDPRHHDY 263
 QY 502 --DRKRHRPVRMTLPCCPD-----MDLMEAKKEQAVFELMRTFKLPGEKINDVMPY 554
 Db 264 INADR-----FIAFLHEIGADAVDHLHVMIEAKKDLALFQLMK-----DLAEY 307
 QY 555 D 555
 Db 308 D 308



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 14:06:28 ; Search time 134.53 seconds
(without alignments)
630.625 Million cell updates/sec

Title: US-09-724-296-38

Perfect score: 3055

Sequence: 1 MGTGLESLSIGDRGAAPTV.....HKLKHYDHLWENHEKLSLS 580

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_17.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	3055	100.0	580	095243	095243 homo sapien
2	1821.5	59.6	554	11 Q9Z2D7	Q9Z2D7 mus musculus
3	880.5	28.8	416	13 Q919F1	Q919F1 gallus gall
4	344	11.3	419	10 Q9SFC1	Q9SFC1 arabidopsis
5	220.5	7.2	467	13 Q9YGC6	Q9YGC6 xenopus lae
6	202.5	6.6	344	13 Q42403	Q42403 gallus gall
7	146	4.8	702	10 Q9SF22	Q9SF22 arabidopsis
8	143.5	4.7	782	5 Q25875	Q25875 plasmodium
9	141.5	4.6	782	5 Q25730	Q25730 plasmodium
10	141.5	4.6	782	5 Q26007	Q26007 plasmodium
11	140.5	4.6	782	5 Q9U430	Q9U430 plasmodium
12	140.5	4.6	782	5 Q9U414	Q9U414 plasmodium
13	138.5	4.5	782	5 Q9U429	Q9U429 plasmodium
14	137.5	4.5	782	5 Q26104	Q26104 plasmodium
15	137.5	4.5	2209	5 Q9U0G6	Q9U0G6 plasmodium
16	136.5	4.5	782	5 Q9U431	Q9U431 plasmodium
17	136.5	4.5	1359	5 O02061	O02061 caenorhabdi
18	135.5	4.4	282	13 Q9PUM9	Q9PUM9 xenopus lae
19	135.5	4.4	755	4 O43719	O43719 homo sapien

20 135.5 4.4 1058 5 09V433
21 135 4.4 822 3 09USH9
22 133.5 4.4 1359 5 09U7E0
23 132.5 4.3 286 1 09YDP0
24 132 4.3 1819 2 09ZLV0
25 131.5 4.3 230 1 09P9L6
26 131.5 4.3 285 11 09Z2D8
27 131.5 4.3 291 4 095983
28 131 4.3 754 4 099730
29 130.5 4.3 411 4 09UBB5
30 130 4.3 537 5 09VZT4
31 130 4.3 1927 2 025262
32 129.5 4.2 3122 11 09JMD6
33 129 4.2 1163 4 09UBH7
34 128.5 4.2 946 10 09C9D8
35 128 4.2 853 10 09LFE4
36 127.5 4.2 1058 5 09U5E0
37 127 4.2 648 10 09LQV6
38 127 4.2 970 10 09AVF2
39 127 4.2 1723 2 09JMX8
40 126.5 4.1 1183 5 09J3D1
41 126 4.1 303 13 09PUM8
42 126 4.1 561 10 09C6S8
43 126 4.1 881 5 09BKN8
44 125.5 4.1 302 4 09UIS8
45 125 4.1 873 11 09WVC9

ALIGNMENTS

RESULT 1
095243
ID 095243 PRELIMINARY; PRT; 580 AA.
AC 095243;
DT 01-MAY-1999 (TREMREL. 10, Created)
DT 01-MAY-1999 (TREMREL. 10, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE METHYL-CPG BINDING PROTEIN MBD4.
GN MBD4 OR MBD1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98449942; PubMed=9774669;
RA Hendrich B., Bird A.
RT Identification and characterization of a family of mammalian methyl-
CPG binding proteins.;
RI Mol. Cell. Biol. 18:6538-6547(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=FETAL BRAIN;
RA MEDLINE=9919294; PubMed=10097147;
RA Bellacosa A., Cicchilitti L., Schepis F., Riccio A., Yeung A.T.,
Matsumoto Y., Gicenis E.A., Genuardi M., Neri G.;
RT "MED1, a novel human methyl-CpG-binding endonuclease, interacts with
DNA mismatch repair protein MLH1".
EL Proc. Natl. Acad. Sci. U.S.A. 96:3969-3974(1999).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=99373255; PubMed=10441743;
RA Hendrich B., Abbott C., McQueen H., Chambers D., Cross S., Bird A.;
RT "Genomic structure and chromosomal mapping of the murine and human
mbd1, mbd2, mbd3, and mbd4 genes".;
RL Mamm. Genome 10:906-912(1999).
DR EMBL; AF072250; AAC68879.1; -
DR EMBL; AF114784; AAD22195.1; -
DR EMBL; AF120999; AAD50374.1; -
DR EMBL; AF120997; AAD50374.1; JOINED.
DR EMBL; AF120998; AAD50374.1; JOINED.
DR Interpro; IPR001739; MBD.



DR InterPro: IPR003265; Endo_3c.
DR Pfam: PF00730; HhH-GPD; 1.
DR Pfam: PF01429; MBD; 1.
DR SMART: SM00391; MBD; 1.
KW Endonuclease.
SQ SEQUENCE 580 AA; 66050 MW; BF16FB21A34B8E5F CRC64;

Query Match 100.0%; Score 3055; DB 4; Length 580;
Best Local Similarity 100.0%; Pred. No. 1.9e-219;
Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGTGLESLSLGRGAAPVTTSERLVPPDPNDLRKEDVAMELERVGEDEEOMIKRSE 60
DB 1 MGTGLESLSLGRGAAPVTTSERLVPPDPNDLRKEDVAMELERVGEDEEOMIKRSE 60
QY 61 CNPLLEPTASAGFATAGTECKRSVPCGWERVVKORLFGKTAGDFDYVFISPOGLKFRS 120
DB 61 CNPLLEPTASAGFATAGTECKRSVPCGWERVVKORLFGKTAGDFDYVFISPOGLKFRS 120
QY 121 KSSLANYLHKNGETSLKPEDFTVLSKRGKIKSRVKDCSMAALTSHLQNSNNNLT 180
DB 121 KSSLANYLHKNGETSLKPEDFTVLSKRGKIKSRVKDCSMAALTSHLQNSNNNLT 180
QY 181 RSKCKKDVMPSPSSSELOESRGLSNFTSTHLLKREDEGVDDVNFRRKPKGKVTILKG 240
DB 181 RSKCKKDVMPSPSSSELOESRGLSNFTSTHLLKREDEGVDDVNFRRKPKGKVTILKG 240
QY 241 IPIKTKKGRKSCSFVQSDSKRESVCNKADAESPVAQKSQLDRTVCISDAGAGETL 300
DB 241 IPIKTKKGRKSCSFVQSDSKRESVCNKADAESPVAQKSQLDRTVCISDAGAGETL 300
QY 301 SVTSENSLVKKERSLSGSGNFCSEQKTSIINKFCSAKDSHNEKYEDTFLESEEIGT 360
DB 301 SVTSENSLVKKERSLSGSGNFCSEQKTSIINKFCSAKDSHNEKYEDTFLESEEIGT 360
QY 361 KVEVERKEHLHTDILKRGSEMDNNSCPTRKDTGKIFQEDTIPRTQIERKTSLYFSS 420
DB 361 KVEVERKEHLHTDILKRGSEMDNNSCPTRKDTGKIFQEDTIPRTQIERKTSLYFSS 420
QY 421 KYNKEALSPRRKAFKKWTTPRSPFNVLQETLFDHPWKLLIATIFLNRSGKMAIPVLWK 480
DB 421 KYNKEALSPRRKAFKKWTTPRSPFNVLQETLFDHPWKLLIATIFLNRSGKMAIPVLWK 480
QY 481 FLEKPSAEVARTADRDVSELLKPLGLYDLRAKTIIVKFSDEYLTQWKYPIELHGIGKY 540
DB 481 FLEKPSAEVARTADRDVSELLKPLGLYDLRAKTIIVKFSDEYLTQWKYPIELHGIGKY 540
QY 541 GNDSYRIFCVNWKQVHPEDHKLKYNKYHDWLWENHEKLSLS 580
DB 541 GNDSYRIFCVNWKQVHPEDHKLKYNKYHDWLWENHEKLSLS 580

RESULT 2

Q92D7 ID Q92D7 PRELIMINARY; PRT; 554 AA.
AC Q92D7;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE METHYL-CPG BINDING PROTEIN MBD4.
GN MBD4.
OS Mus musculus (Mouse), and
OS Mus musculus domesticus (western European house mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090, 10092;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98449942; PubMed=9774669;
RA Hendrich B., Bird A.;
RT "Identification and characterization of a family of mammalian methyl-
CPG binding proteins.";

RL Mol. Cell. Biol. 18:6538-6547(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RX MEDLINE=99373255; PubMed=10441743;
RA Hendrich B., Abbott C., McQueen H., Chambers D., Cross S., Bird A.;
RT "Genomic structure and chromosomal mapping of the murine and human
Mamm. Genome 10:906-912(1999).
RL ENBL; AF120996; AAC68878.1; -;
DR ENBL; AF120996; AAC68878.1; -;
DR MGD; MGI:1333850; Mbd4.
DR InterPro: IPR003265; Endo_3c.
DR InterPro: IPR001739; MBD.
DR InterPro: IPR002052; N6_Mtase.
DR Pfam: PF00730; HhH-GPD; 1.
DR Pfam: PF01429; MBD; 1.
DR SMART: SM00391; MBD; 1.
DR PROSITE: PS00092; N6_Mtase; UNKNOWN_1.
SQ SEQUENCE 554 AA; 62577 MW; 792D37CB180291F5 CRC64;

Query Match 59.6%; Score 1821.5; DB 11; Length 554;
Best Local Similarity 66.2%; Pred. No. 1.4e-127;
Matches 384; Conservative 49; Mismatches 116; Indels 31; Gaps 11;
QY 6 LESLSLGD---RGAAPTTSERLVPPDPNDLRKEDVAMELERVGEDEEOMIKRSECN 62
DB 1 MESPLGDNVRVG-----ESLVPPDPNDLRKEDVAMELERVGEDEEOMIKRSECN 62
QY 63 PLLEPTASAGFATAGTECKRSVPCGWERVVKORLFGKTAGDFDYVFISPOGLKFRSKS 122
DB 51 SLLOEPTAST--LSSTATTGCHKPVPCGWERVVKORLFGKTAGDFDYVFISPOGLKFRSKR 109
QY 123 SLANYLHKNGETSLKPEDFTVLSKRGKIKSRVKDCSMAALTSHLQNSNNNLT 182
DB 110 SLANYLHKNGETSLKPEDFTVLSKRGKIKSRVKDCSMAALTSHLQNSNNNLT 169
QY 183 KCKKDVMPSPSSSELOESRGLSNFTSTHLLKREDEGVDDVNFRRKPKGKVTILKGIP 242
DB 170 KWKTDVLPSPGTSSESSSGLSNSACLLLRHREDIQDVSEKRRKSKRVTLKGTA 229
QY 243 IKTKKGRKSCSFVQSDSKRESVCNKADAESPVAQKSQLDRTVCISDAGAGETLSV 302
DB 230 SOKTKQKCRKSLLESTQNRNRKRSVQKVGADRELVPQESQLNFTLCPADACA--RETGVL 288
QY 303 TSENSLVKKERSLSGSGNFCSEQKTSIINKFCSAKDSHNEKYEDTFLESEEIGTKV 362
DB 289 AGE-----EKSPFGLDLCFIQVTSGTNKPSTEAAGEANR--EQTLESEIRSK - 338
QY 363 EVVERK--EHLHTDILKRGSEMDNNSCPTRKDTGKIFQEDTIPRTQIERKTSLYFSS 420
DB 339 --GDRKGEAHLHTGVLDQSGEMP--SCSQAKHFTSE--TFQEDSIPTQVEKRRKTSLYFSS 394
QY 421 KYNKEALSPRRKAFKKWTTPRSPFNVLQETLFDHPWKLLIATIFLNRSGKMAIPVLWK 480
DB 395 KYNKEALSPRRKAFKKWTTPRSPFNVLQETLFDHPWKLLIATIFLNRSGKMAIPVLWK 454
QY 481 FLEKPSAEVARTADRDVSELLKPLGLYDLRAKTIIVKFSDEYLTQWKYPIELHGIGKY 540
DB 455 FLEKPSAEVARTADRDVSELLKPLGLYDLRAKTIIVKFSDEYLTQWKYPIELHGIGKY 514
QY 541 GNDSYRIFCVNWKQVHPEDHKLKYNKYHDWLWENHEKLSLS 580
DB 515 GNDSYRIFCVNWKQVHPEDHKLKYNKYHDWLWENHEKLSLS 554
RESULT 3
Q919F1 ID Q919F1 PRELIMINARY; PRT; 416 AA.
AC Q919F1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)



Result No.	Score	Query Match %	Length	DB ID	Description
1	3112	71.6	599	2 S71134	UV-endonuclease -
2	1184	27.2	219	2 A28484	glutathione transf
3	1011.5	23.3	656	2 S55262	UV-endonuclease -
4	984	22.6	218	2 A45556	glutathione S-trans
5	921	21.2	218	2 A45523	glutathione transf
6	654	15.0	209	2 A48388	glutathione S-trans
7	481.5	11.1	218	2 S33860	glutathione transf
8	481.5	11.1	218	2 A25794	glutathione transf
9	478.5	11.0	218	2 S32425	glutathione transf
10	477.5	11.0	218	2 S65674	glutathione transf
11	474.5	10.9	218	2 A47486	glutathione transf
12	473.5	10.9	217	2 JX0095	glutathione transf
13	473.5	10.9	218	2 A39375	glutathione transf
14	470	10.8	218	2 A33732	glutathione transf
15	468.5	10.8	218	2 S3202	glutathione transf
16	467.5	10.7	218	2 B34159	glutathione transf
17	465.5	10.7	218	2 S01719	glutathione transf
18	463.5	10.7	218	2 A46048	glutathione transf
19	461.5	10.6	218	2 A29036	glutathione transf
20	460.5	10.6	218	2 B92331	glutathione transf
21	455.5	10.5	218	1 X0RT04	glutathione transf
22	454	10.4	218	2 A46143	glutathione transf
23	452.5	10.4	218	2 B28946	glutathione transf
24	447.5	10.3	225	2 A35295	mu-class glutathio
25	429.5	9.9	220	2 S18464	glutathione transf
26	385.5	8.9	219	2 S50146	glutathione transf
27	360	8.3	320	2 S55418	glutathione transf
28	306.5	7.0	322	2 S83974	major allergen Dpl
29	257.5	5.9	142	2 S17462	UV-endonuclease ho hypothetical prote glutathione transf

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

9, T. 6k20.

Tue Jan 15 14:12:06 2002

Db 241 DKPIPWGRGLYACINTILSRMKRERFVCSRTCTRTTTIQRDGLSVKOLGTQNVLDLIKLV 300
 QY 530 EWNHNFGLHFRVNSDLPFPFASHAKYGYTLEFAQSHLEEVGKLANKNHRLTHMHPGGYTQ 589
 Db 301 EWNHNFGLHFRVNSDLPFPFASHAKYGYTLEFAQSHLEEVGKLANKNHRLTHMHPGGYTQ 360
 QY 590 IASPREVVVDSAIRDLAYHDEILSRMKLNEQLNKDAVLIHLGTFEGKKTETLDRFRKNY 649
 Db 361 IASPREVVVDSAIRDLAYHDEILSRMKLNEQLNKDAVLIHLGTFEGKKTETLDRFRKNY 420
 QY 650 QRLSDSVKARLVLENDVSWSDLLPLCQELNPLVLDWHHNIIVPGLTREGSLDMLPL 709
 Db 421 QRLSDSVKARLVLENDVSWSDLLPLCQELNPLVLDWHHNIIVPGLTREGSLDMLPL 480
 QY 710 IPTIRETWRTKGITOKOHYSADPTAISGMKRRRAHSDRVDFPCCDPTMDLMTAEKE 769
 Db 481 IPTIRETWRTKGITOKOHYSADPTAISGMKRRRAHSDRVDFPCCDPTMDLMTAEKE 540
 QY 770 QAVFELCRYELQNPCCPLEMGPEYDQTRDGYPPGAEKRLTARKRSRKEEVEDEK 828
 Db 541 QAVFELCRYELQNPCCPLEMGPEYDQTRDGYPPGAEKRLTARKRSRKEEVEDEK 599

RESULT 2

A26484
 glutathione transferase (EC 2.5.1.18) - fluke (Schistosoma japonicum) (fragment)
 C:Species: Schistosoma japonicum
 C>Date: 25-Oct-1987 #sequence_revision 30-Sep-1989 #text_change 03-Feb-1994
 C:Accession: A94139; A26484; A28315
 R:Smith, D.B.; Davern, K.M.; Board, P.G.; Tiu, W.U.; Garcia, E.G.; Mitchell, G.F.
 Proc. Natl. Acad. Sci. U.S.A. 83, 8703-8707, 1986
 A:Title: Mr 26,000 antigen of Schistosoma japonicum recognized by resistant WEHI 129/J m
 A:Reference number: A94139; MUID:87041520
 A:Accession: A94139
 A:Molecule type: mRNA
 A:Residues: 1-219 <SMI>
 R:Smith, D.B.; Davern, K.M.; Board, P.G.; Tiu, W.U.; Garcia, E.G.; Mitchell, G.F.
 Proc. Natl. Acad. Sci. U.S.A. 84, 6541, 1987
 A:Reference number: A94181
 A:Contents: annotation; revision to residues 210-219
 C:Superfamily: glutathione transferase
 C:Keywords: transferase

Query Match 27.2%; Score 1184; DB 2; Length 219;
 Best Local Similarity 100.0%; Pred. No. 6.7e-58;
 Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 TKLPILGYWKIKGLVOPTRLLLEYLEEKYEHLRYERDEGDKWRNKKFELGLFPNLPYYI 61
 Db 1 TKLPILGYWKIKGLVOPTRLLLEYLEEKYEHLRYERDEGDKWRNKKFELGLFPNLPYYI 60
 QY 62 DGDVKLTQSAIIRYIADKNHMLGGCPKRAETSMLEGAVIDIRYGVSRVIAYSKDFETLK 121
 Db 61 DGDVKLTQSAIIRYIADKNHMLGGCPKRAETSMLEGAVIDIRYGVSRVIAYSKDFETLK 120
 QY 122 VDFLSKLPMLKMFEDRLCHTKYINGDHVTHPDFMLYDALDVLVYMDPCLDAFPKLVCF 181
 Db 121 VDFLSKLPMLKMFEDRLCHTKYINGDHVTHPDFMLYDALDVLVYMDPCLDAFPKLVCF 180
 QY 182 KKREIATPQIDKYLKSSKYIAPLQWQATFGGDDHPK 220
 Db 181 KKREIATPQIDKYLKSSKYIAPLQWQATFGGDDHPK 219

RESULT 3

S55262
 UV-endonuclease - Neurospora crassa
 C:Species: Neurospora crassa
 C>Date: 10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 20-Jun-2000
 C:Accession: S55262
 R:Yajima, H.; Takao, M.; Yasuhira, S.; Zhao, J.H.; Ishii, C.; Inoue, H.; Yasui, A.
 EMBO J. 14, 2393-2399, 1995

A:Title: A eukaryotic gene encoding an endonuclease that specifically repairs DNA dam
 A:Reference number: S55262; MUID:95292980
 A:Accession: S55262
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-656 <YAJ>
 A:Cross-references: EMBL:D11392; NID:g1526560; PID:g927215
 A:Genetics:
 A:Gene: uvel

Query Match 23.3%; Score 1011.5; DB 2; Length 656;
 Best Local Similarity 36.6%; Pred. No. 7.6e-48;
 Matches 241; Conservative 88; Mismatches 149; Indels 181; Gaps 18;

QY 339 ARSKKKATDSVDKIDESVASYDSST-----HLRSSR-----SKKPVNNSSES- 386
 Db 3 SRKSKAALDTPQSESTFSSTLDSSAPSPARNLRKSGRNILQPSSEKDRDHEKSGEEL 62
 QY 387 -----SEQISKATKKVKQKEEYVEEVEDEK 413
 Db 63 AGRMGKDANGHCLRECKEQEGVKMAIEGLARMERRLQRA TKRQKKOLED----- 114
 QY 414 SLKNESSEDEFEPVPEQLETPISKRRS-----RSAXNLEKESTMNLDHA----- 461
 Db 115 GIPVPSVVSFR-PTADYHHKSTNAEEREAKPEVLKTHSKDVEREAIEGVDVVKMEPAAT 173
 QY 462 -----PREMFCLDK-----PIWGRGLYACINTILSRMKRERFVCSRTCTIT 504
 Db 174 NIEPEDAQDAEAGARPPAVNSVYLPWPWGLRGYACINTYLRNSKPPIFSRTCRMA 233
 QY 505 TI-----ORDGLESVKOLGTQNVLDLIKLVENHNFG 536
 Db 234 STVDHRHPLOFEDPEPHHLKKNPKDKSKEPQDELGHKFPVQELGLANARDIVKMLCWNKYG 293
 QY 537 IHMRVSSDLFPFASHAKYGYTL-EFAQSHLEEVGKLANKNHRLTHMHPGGYTQIASPRE 595
 Db 294 IRLRLSSEMPFPFASHPVHGYKLAPFASEVLAAGRVAAELGHLRTHPQQFTQLGSPRK 353
 QY 596 VVDSAIRDLAYHDEILSRMKLNEQLNKDAVLIHLGTFEGKKTETLDRFRKNYQRLSDS 655
 Db 354 EVESAIRDLAYHDEILSRMKLNEQLNKDAVLIHLGTFEGKKTETLDRFRKNYQRLSDS 413
 QY 656 VKARLVLENDVSWSDLLPLCQELNPLVLDWHHNIIV-PCTLRGSLDLM--PLIP 711
 Db 414 CKNLRLVLENDVGVTVHDLPLVCEELNIPVLDYHHNICFDPANHLREGTLDLSDPKLQE 473
 QY 712 TIRETWRTKGITOKOHYSADPTAISGMKRRRAHSDRVDFPCCDPTMDLMTAEKEQA 771
 Db 474 RIANTWRRKGIKQKMHYSEPCD-CAVTPRHRKRRPRVMTLPPCPDMDLMTAEKDEQA 532
 QY 772 VFELCRYEL-----QNPPCP-----LEIMGPE--- 794
 Db 533 VFELMKTEKLPCEKINDMVPYDRDNRPAVPKPKKKGGKRRKRTTDEAAEPEVD 592
 QY 795 --YQTRDG-----YYPGAGKRLTARKRRSRK-----EEVEED 826
 Db 593 TAADDVKDAPEGPKPEVPEERAMGPNRVVPLGCEENLKPKKREVKKGKVPVEDE 651

RESULT 4

A45556
 glutathione S-transferase - fluke (Schistosoma mansoni)
 C:Species: Schistosoma mansoni
 C>Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 03-May-1996
 C:Accession: A45556
 R:Wright, M.D.; Harrison, R.A.; Melder, A.M.; Newport, G.R.; Mitchell, G.F.
 Mol. Biochem. Parasitol. 49, 177-179, 1991
 A:Title: Another 26-kilodalton glutathione S-transferase of Schistosoma mansoni.
 A:Reference number: A45556; MUID:92131046
 A:Accession: A45556
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid

0
4
4

2
4
4
9
8
9

L.

;**Molecule type:** mRNA

Db 124 KPEFLKTIPEKMKLYSEFLGKRPWFGAGKVTYVDFLAYDILDQYHIEPKCLDAFPNLDK 183
 QY 181 FKKRTAIPQIDKYLKSSKIYAWPL 205
 Db 184 FLARFGLKKISAYMKSSRYLSTPI 208

RESULT 9
 S32425
 glutathione transferase (EC 2.5.1.18) class mu, GSTM4 (version 2) - human
 N:Alternate names: glutathione transferase mu4
 C:Species: Homo sapiens (man)
 C:Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 18-Jun-1999
 C:Accession: S32425; S29337
 R:Zhong, S.; Spurr, N.K.; Hayes, J.D.; Wolf, C.R.
 Biochem. J. 291, 41-50, 1993
 A:Title: Deduced amino acid sequence, gene structure and chromosomal location of a novel
 A:Reference number: S32424; MUID:93228631
 A:Accession: S32425
 A:Molecule type: DNA
 A:Residues: 1-218 <2HO>
 A:Cross-references: EMBL:X68677; NID:g31934; PIDN:CAA48637.1; PID:g31935
 A:Note: The authors translated the codon AAG for residue 182 as Arg and CCA for residue
 C:Genetics:
 A:Gene: GDB:GSTM4
 A:Cross-references: GDB:134191; OMIM:138333
 A:Map position: lp13.3-lp13.3
 A:Introns: 12/3; 38/1; 59/3; 87/1; 120/3; 152/3; 189/3
 C:Superfamily: glutathione transferase
 C:Keywords: dimer; transferase

Query Match 11.0%; Score 478.5; DB 2; Length 218;
 Best Local Similarity 44.1%; Pred. No. 2e-19;
 Matches 94; Conservative 37; Mismatches 77; Indels 5; Gaps 1;
 QY 6 ILGYWIKGLVQPTRLLEYLEEYEEHLYERD-----ECDKWRNKKFELGLEFNPYPY 60
 Db 4 ILGYWDIRGLAHIRLLELVETSSYEKKYTMGGAPDYDRSQWLNEKPKLGLDFNLPYL 63
 QY 61 IDGDVKLTQSMATIRYIADKHNLMGCPKERAIEISMLEGAVLDIRYGVSRVAYSKDFETPL 120
 Db 64 IDGAHITQSNALICVIAKHNLCGGTEEEKIRVDILENQAMDVSNQLARVCYSPDFEKL 123
 QY 121 KVDLSKLPMLKMFEDRLCHKTYLNGDHTVHPDFMLYDALDVLYMDPMCLDAFPKLVLC 180
 Db 124 KPEYLEELPTMQHFSQFLGKRPWFVGDKITFVDFLAYDVLDLHRIEFPNCLDAFPNLDK 183
 QY 181 FKKRIEAIPOIDKYLKSSKIYAWPLOGWOATFG 213
 Db 184 FTSRFGLEKISAYMKSSRFLPKPLYTRMAVWG 216

RESULT 10
 S65674
 glutathione transferase (EC 2.5.1.18) class mu chain 7.8 - rabbit
 C:Species: Oryctolagus cuniculus (domestic rabbit)
 C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 18-Jun-1999
 C:Accession: S65674; S30380
 R:Lee, S.H.; Lee, Y.S.; Han, J.S.; Kim, Y.S.; Koh, J.K.
 Arch. Biochem. Biophys. 318, 424-429, 1995
 A:Title: Cloning and expression of a cDNA for Mu-Class Glutathione S-transferase from ra
 A:Reference number: S65674; MUID:95251394
 A:Accession: S65674
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-218 <LEP>
 A:Cross-references: EMBL:L23766; NID:g388198; PIDN:AAA69665.1; PID:g388199
 R:Primiano, T.; Novak, R.F.
 Arch. Biochem. Biophys. 301, 404-410, 1993
 A:Title: Purification and characterization of class mu glutathione S-transferase isozyme
 A:Reference number: S30380; MUID:93213177
 A:Accession: S30380

A:Molecule type: protein
 A:Residues: 2-21 <PRI>
 C:Superfamily: glutathione transferase
 C:Keywords: dimer; transferase

Query Match 11.0%; Score 477.5; DB 2; Length 218;
 Best Local Similarity 44.6%; Pred. No. 2.3e-19;
 Matches 95; Conservative 36; Mismatches 77; Indels 5; Gaps 1;

QY 7 LGVWKIKGLVQPTRLLEYLEEYEEHLYERDEG-----DKWRNKKFELGLEFNPYPY 61
 Db 5 LGTDVVRGLALPIRMILEYTDTSYEKKYTMGDAPNYDQSKWLSEKFTLGLDFNLPVLI 64
 QY 62 DGDVKLTQSMATIRYIADKHNLMGCPKERAIEISMLEGAVLDIRYGVSRVAYSKDFETLK 121
 Db 65 DGTNKLTSQNALIRYLARKHGLCGTEEEKIRVDILENQMDNRFLQVNVVCYSPDFEKLK 124
 QY 122 VDFLSKLPMLKMFEDRLCHKTYLNGDHTVHPDFMLYDALDVLYMDPMCLDAFPKLVCF 181
 Db 125 PEYLGKPEKQLQYQFLGSLPWFAGDKITFADFVLVDVLDQNRIFVPGCLDAFPNLDKF 184
 QY 182 KKRIEAIPOIDKYLKSSKIYAWPLOGWOATFG 214
 Db 185 HVRFGLPKISAYMKSSRFRIRVPVFLKKATWVG 217

RESULT 11
 A47486
 glutathione transferase (EC 2.5.1.18) class mu, GSTM4 (version 1) - human
 C:Species: Homo sapiens (man)
 C:Date: 02-Jun-1995 #sequence_revision 02-Jun-1995 #text_change 18-Jun-1999
 C:Accession: A47486; B47486; S36782; I37438; S45685
 R:Comstock, K.E.; Johnson, K.J.; Rifkenbery, D.; Henner, W.D.
 J. Biol. Chem. 268, 16958-16965, 1993
 A:Title: Isolation and analysis of the gene and cDNA for a human mu class glutathione
 A:Reference number: A47486; MUID:93352467
 A:Accession: A47486
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-218 <COM>
 A:Cross-references: GB:M96233
 A:Accession: B47486
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-218 <CO2>
 A:Cross-references: GB:M96234; NID:g306818; PIDN:AAA57347.1; PID:g306819
 A:Experimental source: HeLa cells
 R:Ross, V.L.; Board, P.G.
 Biochem. J. 294, 373-380, 1993
 A:Title: Molecular cloning and heterologous expression of an alternatively spliced hu
 A:Reference number: S36782; MUID:93384505
 A:Accession: S36782
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-218 <ROS>
 A:Cross-references: EMBL:M99422
 A:Experimental source: testis
 R:Comstock, K.E.; Widersten, M.; Hao, X.Y.; Henner, W.D.; Mannervik, B.
 Arch. Biochem. Biophys. 311, 487-495, 1994
 A:Title: A comparison of the enzymatic and physicochemical properties of human glutat
 A:Reference number: S45684; MUID:94263230
 A:Contents: annotation
 R:Taylor, J.B.; Oliver, J.; Sherrington, R.; Pemble, S.E.
 Biochem. J. 274, 587-593, 1991
 A:Title: Structure of human glutathione S-transferase class Mu genes.
 A:Reference number: I37438; MUID:91174774
 A:Accession: I37438
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 39-120 <RES>
 A:Cross-references: EMBL:X56837; NID:g31936; PIDN:CAA40167.1; PID:g31937
 C:Genetics:

Tue Jan 15 14:12:06 2002

C:Species: Homo sapiens (man)
 C:Date: 20-Mar-1992; #sequence_revision 06-Jan-1995; #text_change 21-Jul-2000
 A:Cross-references: GDB:134191; OMIM:138333
 A:Map position: 1p13.3-lp13.3
 A:Introns: 12/3; 38/1; 59/3; 87/1; 120/3; 152/3; 189/3
 C:Superfamily: glutathione transferase
 C:Keywords: alternative splicing; dimer; transferase

Query Match 10.9%; Score 474.5; DB 2; Length 218;
 Best Local Similarity 43.9%; Pred. No. 3.4e-19;
 Matches 93; Conservative 37; Mismatches 77; Indels 5; Gaps 1;

QY 7 LGYWKIKGLVQPTLLLEYLEEKYEHEHYERDEG-----DKWRNKKFGLGFPNLPYI 61
 ||||| ||| ||||| ||| ||| : ||||| ||||| |||
 Db 5 LGYWDIRGLAHAIIRLLLEVTDSSEYEEKYTMGDAPDYDRSQWLNKFKLGLDFPNLPYLI 64
 ||||| ||| ||||| ||| ||| : ||||| ||||| |||
 QY 62 DGDVKTQSMALIRYIADKHNLGCGPKRAEISMLEGAVLDIRYGVSRVIAYSKDFETLK 121
 ||||| ||| ||||| ||| ||| : ||||| ||||| |||
 Db 65 DGAHKITQSNALICYIARKHNLCGETEEKIRVDILENQAMDVSQNLARVCYSPDFEKLK 124
 ||||| ||| ||||| ||| ||| : ||||| ||||| |||
 QY 122 VDFLSKLPKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVLYMDPCLDAPFKLVCF 181
 ||||| ||| ||||| ||| ||| : ||||| ||||| |||
 Db 125 PEYLELPTMQHFSQFLGKRFWFGDKITTFVDFLAYDVLDRHIFEPNCLDAPFNKDF 184
 ||||| ||| ||||| ||| ||| : ||||| ||||| |||
 QY 182 KKRIEAIPQIDKYLKSSKYIAMPLOQWATFG 213
 ||||| ||| ||||| ||| ||| : ||||| ||||| |||
 Db 185 ISRFEGLEKISAYKSSRFLPKPLYTRVAVWG 216
 ||||| ||| ||||| ||| ||| : ||||| ||||| |||

RESULT 12

JX0095
 glutathione transferase (EC 2.5.1.18) b - guinea pig
 N:Alternate names: glutathione S-alkyltransferase; glutathione S-aryltransferase; glutathione transferase (guinea pig)
 C:Species: Cavia porcellus (guinea pig)
 C:Date: 07-Sep-1990; #sequence_revision 07-Sep-1990; #text_change 18-Jun-1993
 C:Accession: JX0095
 R:Kamel, K.; Oshino, R.; Hara, S.
 J. Biochem. 107, 111-117, 1990
 A:Title: Amino acid sequence of glutathione S-transferase b from guinea pig liver.
 A:Reference number: JX0095; MUID:90236961
 A:Accession: JX0095
 A:Molecule type: protein
 A:Residues: 1-217 <KAM>
 C:Comment: Glutathione transferases are a family of enzymes that are multifunctional in
 C:Superfamily: glutathione transferase
 C:Keywords: transferase

Query Match 10.9%; Score 473.5; DB 2; Length 217;
 Best Local Similarity 45.1%; Pred. No. 3.8e-19;
 Matches 92; Conservative 34; Mismatches 73; Indels 5; Gaps 1;

QY 7 LGYWKIKGLVQPTLLLEYLEEKYEHEHYERDEG-----DKWRNKKFGLGFPNLPYI 61
 ||||| ||| ||||| ||| ||| : ||||| ||||| |||
 Db 4 LGYWNIRGLTHPIRLILEYNTSGYEKKRYNMGDAPDYDRSQWLNKFKLGLDFPNLPYLI 63
 ||||| ||| ||||| ||| ||| : ||||| ||||| |||
 QY 62 DGDVKTQSMALIRYIADKHNLGCGPKRAEISMLEGAVLDIRYGVSRVIAYSKDFETLK 121
 ||||| ||| ||||| ||| ||| : ||||| ||||| |||
 Db 64 DGTGHLKTSNAILRYIARKHNLCGVTEETIRMDILENQVMDIRMDILMICYSPDFEQK 123
 ||||| ||| ||||| ||| ||| : ||||| ||||| |||
 QY 122 VDFLSKLPKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVLYMDPCLDAPFKLVCF 191
 ||||| ||| ||||| ||| ||| : ||||| ||||| |||
 Db 124 AEFLEGIPDKMKLFQFLGKLPWFAGNKLTYVDFLAYDVLDRHIFEPNCLDAPFNKDF 183
 ||||| ||| ||||| ||| ||| : ||||| ||||| |||
 QY 182 KKRIEAIPQIDKYLKSSKYIAMPLOQWATFG 205
 ||||| ||| ||||| ||| ||| : ||||| ||||| |||
 Db 184 ISRFEGLEKISAYKSSRFLPKPL 207
 ||||| ||| ||||| ||| ||| : ||||| ||||| |||

RESULT 13

A39375
 glutathione transferase (EC 2.5.1.18) mu-2 [validated] - human
 N:Alternate names: glutathione S-transferase GST4, muscle; glutathione transferase GST6;

Query Match 10.9%; Score 473.5; DB 2; Length 218;
 Best Local Similarity 44.3%; Pred. No. 3.8e-19;
 Matches 94; Conservative 35; Mismatches 78; Indels 5; Gaps 1;

QY 7 LGYWKIKGLVQPTLLLEYLEEKYEHEHYERDEG-----DKWRNKKFGLGFPNLPYI 61
 ||||| ||| ||||| ||| ||| : ||||| ||||| |||
 Db 5 LGYWNIRGLTHPIRLILEYNTSGYEKKRYNMGDAPDYDRSQWLNKFKLGLDFPNLPYLI 64
 ||||| ||| ||||| ||| ||| : ||||| ||||| |||
 QY 62 DGDVKTQSMALIRYIADKHNLGCGPKRAEISMLEGAVLDIRYGVSRVIAYSKDFETLK 121
 ||||| ||| ||||| ||| ||| : ||||| ||||| |||
 Db 65 DGTGHLKTSNAILRYIARKHNLCGVTEETIRMDILENQVMDIRMDILMICYSPDFEQK 124
 ||||| ||| ||||| ||| ||| : ||||| ||||| |||
 QY 122 VDFLSKLPKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVLYMDPCLDAPFKLVCF 181
 ||||| ||| ||||| ||| ||| : ||||| ||||| |||
 Db 125 PEYQLAPKMLKLYSQFLGKQPFWFGDKITTFVDFLAYDVLDRHIFEPNCLDAPFNKDF 184
 ||||| ||| ||||| ||| ||| : ||||| ||||| |||
 QY 182 KKRIEAIPQIDKYLKSSKYIAMPLOQWATFG 213
 ||||| ||| ||||| ||| ||| : ||||| ||||| |||

db 185 ISRFEGLEKISAYMKSSRFLPRPVFTKMAVWG 216

RESULT 14

A23732
glutathione transferase (EC 2.5.1.18) mu - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C:Date: 30-Dec-1991 #sequence_revision 30-Dec-1991 #text_change 18-Jun-1999
C:Accession: A23732; S21908
R:Norris, J.S.; Schwartz, D.A.; MacLeod, S.L.; Fan, W.; O'Brien, T.J.; Harris, S.E.; Timpl
Mol. Endocrinol. 5, 979-986, 1991
A:Title: Cloning of a mu-class glutathione S-transferase complementary DNA and character
A:Reference number: A23732; MUID:92049380
A:Accession: A23732
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-218 <NOR>
A:Cross-references: EMBL:X61033; NID:g49638; PIDN:CAA43368.1; PTD:g49639
A:Note: submitted to the EMBL Data Library, July 1991
C:Superfamily: glutathione transferase
C:Keywords: transferase

Query Match 10.8%; Score 470; DB 2; Length 218;
Best Local Similarity 44.3%; Pred. No. 5.9e-19;
Matches 90; Conservative 39; Mismatches 68; Indels 6; Caps 2;

QY	4	LPI-IGWYKIGLVQPRRLLEYLEEKYEHHLYERDEG----	DKNRKKFKELGLEFPNL	57
Db	1	MPVTIGLYWDIRGLAHAIIRLLLELTDTSYEEKYTWGDAPN	FDRSQWLNEKFKLGJDFPNL	60
QY	58	PYYIDGDKVLTQSMARIYTADKHNMJGCCPKERAIEISL	EGAVLDITRYGVSRIATSKDF	117
Db	61	PYLIDGSHKITQSNAILRYIARKHJCGETEERTQLDILEN	QAMDTRMQLAMVCYSPDF	120
QY	118	ETLKVDFSLPEMLKMFEDRILCHTYLNGDRVTHVDFEML	YDALDVLVYDMPCLDAEPK	177
Db	121	EKRPEYLEGLPEKMKLYSEFLGKRSWFAGDKITTVDFLI	YDVLVDQHRIFAPKCLDAFPN	180
QY	178	LVCFKRIEATPOIDKYLSSKY	200	
Db	181	LKDFLARFEGLKIKISDMKSSRF	203	

RESULT 15

SI3202
glutathione transferase (EC 2.5.1.18) Y1 - Chinese hamster
C:Species: Cricetulus griseus (Chinese hamster)
C:Date: 18-Feb-1994 #sequence_revision 06-Jan-1995 #text_change 18-Jun-1999
C:Accession: SI3202
R:Robert de Saint Vincent, B.; Hyrien, O.; Debatisse, M.; Buttin, G.
E:R. J. Biochem. 193, 19-24, 1990
A:Title: Complication of mu class glutathione S-transferase genes and an adenylate de
A:Reference number: SI3202; MUID:91031445
A:Accession: SI3202
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-218 <R0>
A:Cross-references: GB:X57489; MID:g49541; PIDN:CAA40726.1; PID:g49542
A:Note: in Genbank entry CLY1, release 109, the source is designated as Cricetulus longi
C:Superfamily: glutathione transferase
C:Keywords: dimer; transferase

Query Match 10.8%; Score 468.5; DB 2; Length 218;
Best Local Similarity 43.4%; Pred. No. 7.2e-19;
Matches 89; Conservative 36; Mismatches 75; Indels 5; Gaps 1;

6 ILGWYKIGLVQPTRLLEYLEEKYTEHLVERDEG----DKWRNKKFELGLEFPNLPYY 60
 |||| :|| | ||||| : ||| | : ||||| : |||||
 4 ILGWYNYRGLTNPTRLLEYLTDSSYEKKYTMGDAPDSRQWLNEKFKLGIDFPNLPYL 63
 61 IDGVKUTQSMAIRYTADKHNMLGGCPKRAEISMLEGAVLDIRYGSRIAYSKDFETL 120

Db	64	IDGSHKITSNALRYIARKHNLCGTEEBERKVDIVENQAMDRMQLIMLCYNPDFEQ	123
QY	121	KVDFLSKLPMLKNMFEDRLCHKTYLNGDIHVTHPDFMLVDALDVLVYMDPCLIDAFPKLVC	180
Db	124	KPEFLKTIPEKKMKMYSEFLCKRPWFASDKVYLCGFLAYDVLDOYQMFEPKCLDPFNNLKD	183
QY	181	FKYRIEAIPOIDKYLKSSYIAWPL	205
Db	184	FLARFECLKKISAYMKTSRFLRPI	208

Search completed: January 15, 2002, 13:52:18
Job time: 387 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 14:06:15 ; Search time 134.53 Seconds
(without alignments)
713.258 Million cell updates/sec

Title: US-09-724-296-36

Perfect score: 3482

Sequence: 1 MPSRKSKAAALDTPQESST.....REVKKGVPEVEDEGEFDG 656

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_17.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	3482	100.0	656	Q01408	Q01408 neurospora
2	1011.5	29.0	599	Q10988	Q10988 schizosacch
3	377	10.8	322	Q9K9P8	Q9K9P8 bacillus ha
4	359	10.3	317	Q9L4F0	Q9L4F0 bacillus ce
5	359	10.3	325	Q9L4E9	Q9L4E9 bacillus ce
6	261.5	7.5	305	Q9S0M9	Q9S0M9 deinococcus
7	261.5	7.5	326	Q9RTE6	Q9RTE6 deinococcus
8	160.5	4.6	1280	Q75158	Q75158 homo sapien
9	160.5	4.6	1341	Q9UKV3	Q9UKV3 homo sapien
10	155.5	4.5	5327	Q76891	Q76891 drosophila
11	152.5	4.4	1763	Q9JKX5	Q9JKX5 mus musculu
12	144	4.1	16215	Q9NFS3	Q9NFS3 drosophila
13	143.5	4.1	828	Q9VEU8	Q9VEU8 drosophila
14	143.5	4.1	914	Q9H945	Q9H945 homo sapien
15	141.5	4.1	1110	Q91255	Q91255 petromyzon
16	141	4.0	776	Q9H1R4	Q9H1R4 homo sapien
17	140.5	4.0	1338	Q9JIX8	Q9JIX8 mus musculu
18	140.5	4.0	3263	Q917U3	Q917U3 drosophila
19	140.5	4.0	6815	Q917U4	Q917U4 drosophila

20	139.5	4.0	1510	5	061802	061802 caenorhabdi
21	139	4.0	1815	2	Q9K0B4	Q9K0B4 neisseria m
22	138.5	4.0	1520	4	015087	015087 homo sapien
23	138.5	4.0	1781	4	Q9UKX0	Q9UKX0 homo sapien
24	138	4.0	1547	5	Q26471	Q26471 schistocerc
25	137.5	3.9	806	10	Q9M8T5	Q9M8T5 arabidopsi
26	137.5	3.9	964	5	Q9VKD8	Q9VKD8 drosophila
27	137.5	3.9	1197	3	Q9C3Y7	Q9C3Y7 candida alb
28	137	3.9	735	5	Q9VTN4	Q9VTN4 drosophila
29	137	3.9	785	5	Q9GQ82	Q9GQ82 drosophila
30	136.5	3.9	1591	11	P97868	P97868 mus musculu
31	136	3.9	3261	4	Q9V556	Q9V556 homo sapien
32	136	3.9	3942	11	088737	088737 mus musculu
33	135.5	3.9	971	5	045785	045785 caenorhabdi
34	135	3.9	767	5	Q9U234	Q9U234 caenorhabdi
35	135	3.9	1530	11	Q9EQZ7	Q9EQZ7 mus musculu
36	134.5	3.9	837	5	Q9VYV7	Q9VYV7 drosophila
37	134.5	3.9	1695	5	Q9NK53	Q9NK53 drosophila
38	134.5	3.9	1711	5	Q9VJL0	Q9VJL0 drosophila
39	134	3.8	3576	11	Q9QZK2	Q9QZK2 mus musculu
40	133.5	3.8	583	4	Q9UKV2	Q9UKV2 homo sapien
41	133.5	3.8	980	5	Q96246	Q96246 plasmodium
42	133.5	3.8	1399	11	Q9JIS0	Q9JIS0 rattus norv
43	133	3.8	718	13	073619	073619 xenopus lae
44	133	3.8	795	5	Q9V8K6	Q9V8K6 drosophila
45	133	3.8	2073	4	Q9UKW2	Q9UKW2 homo sapien

ALIGNMENTS

RESULT 1
Q01408
ID Q01408
AC Q01408; PRELIMINARY: PRT: 656 AA.
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE UV-ENDORNUCLEASE.
GN UVEL.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A;
RX MEDLINE=95292980; PubMed=7774597;
RA Yajima H., Takao M., Yasuhira S., Zhao J.H., Ishii C., Inoue H.,
RA Yasui A.;
RT "A eukaryotic gene encoding an endonuclease that specifically repairs
RT DNA damaged by ultraviolet light.";
RL EMBO J. 14:2393-2399(1995).
DR EMBL; D11392; CAB20113.1; -.
KW Endonuclease.
SQ SEQUENCE 656 AA; 74413 MW; 4B231E7A8735C86F CRC64;

Query Match	100.0%	Score 3482;	DB 3;	Length 656;
Best Local Similarity	100.0%;	Pred. No. 4.4e-227;		
Matches 656;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MPSRKSKAAALDTPQESSTFSSTLDSAPSAPNLRSGRNILQPSSEKDRDHEKRSGE 60		
Db	1	MPSRKSKAAALDTPQESSTFSSTLDSAPSAPNLRSGRNILQPSSEKDRDHEKRSGE 60		
Qy	61	ELAGRMGKDGANGHCLREGKEQEGVKMAIEGLARMPERLQRTATKROKLEEDGIPVPS 120		
Db	61	ELAGRMGKDGANGHCLREGKEQEGVKMAIEGLARMPERLQRTATKROKLEEDGIPVPS 120		
Qy	121	VVSREPTAPYHHKSTNAEREAKEPVLKTHSKOVREAEIGVDVVKMEPAATNIIEPED 180		
Db	121	VVSREPTAPYHHKSTNAEREAKEPVLKTHSKOVREAEIGVDVVKMEPAATNIIEPED 180		

```

QY 181 AQDAERGAARPPAVNSSLPLPWKGRIGYACLTLYLRNAKPPPIFSRTRCMAASIVDHRH 240
DB 181 AQDAERGAARPPAVNSSLPLPWKGRIGYACLTLYLRNAKPPPIFSRTRCMAASIVDHRH 240
QY 241 PLOFDEPEHLLKNNPKDSKEPQDELGHKGFVOELGLANARDIVKMLCWNKYGIRFRLS 300
DB 241 PLOFDEPEHLLKNNPKDSKEPQDELGHKGFVOELGLANARDIVKMLCWNKYGIRFRLS 300
QY 301 SEMEPFASHPVHGYKLAPFASVLAELAGRAVAELGHLRTHHPGQFTOLGSPKRVESAI 360
DB 301 SEMEPFASHPVHGYKLAPFASVLAELAGRAVAELGHLRTHHPGQFTOLGSPKRVESAI 360
QY 361 RDLHYHDELLSLLKLPEQONRDVMIHMGQFGDKAATLRFKRNRYARLSOSCKNRLVL 420
DB 361 RDLHYHDELLSLLKLPEQONRDVMIHMGQFGDKAATLRFKRNRYARLSOSCKNRLVL 420
QY 421 ENDDVGVTHDILLPVCEELNPMVLDYHHNICFDPHAHLREGTLDISDPKLOERIANTWK 480
DB 421 ENDDVGVTHDILLPVCEELNPMVLDYHHNICFDPHAHLREGTLDISDPKLOERIANTWK 480
QY 481 RKGIKOKMHYSEPCDGAVTPRDRKRHRPRVMTLPPCPDMDLMIEAKDKQAVFELMRTF 540
DB 481 RKGIKOKMHYSEPCDGAVTPRDRKRHRPRVMTLPPCPDMDLMIEAKDKQAVFELMRTF 540
QY 541 KLPFGFEKINDMVPYDRDDENRPPVAPKPKKGGKRRKTTDEEAAPEEVDTAADDVKD 600
DB 541 KLPFGFEKINDMVPYDRDDENRPPVAPKPKKGGKRRKTTDEEAAPEEVDTAADDVKD 600
QY 601 APEGKPEVEERAMGGPNRYWPLGCEWLKPKKREVKKGKVPVEEDSGEDFG 656
DB 601 APEGKPEVEERAMGGPNRYWPLGCEWLKPKKREVKKGKVPVEEDSGEDFG 656

RESULT 2
ID Q10988 PRELIMINARY; PRT; 599 AA.
AC Q10988; P87339;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE UV-ENDORNUCLEASE.
GN UVDE OR UVEI+.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sp972;
RX Takao M., Yonemasa R., Yamamoto K., Yasui A.;
RA MEDLINE=96188860; PubMed=8614629;
RT "Characterization of a uv endonuclease gene from the fission yeast
RT Schizosaccharomyces pombe and its bacterial homolog.";
RL Nucleic Acids Res. 24:1267-1271(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97175806; PubMed=9023111;
RA Davey S., Nass M.L., Ferrer J.V., Sidik K., Eisenberger A.,
RA Mitchell D.L., Freyer G.A.;
RT "The fission yeast UVDR DNA repair pathway is inducible.";
RL Nucleic Acids Res. 25:1002-1008(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Lyne M., Rajandream M.A., Barrell B.G., Lucas M., Gaillardin C.;
RA Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; D78571; BAAL1415.1; -
DR EMBL; U78487; AAC49664.1; -
DR EMBL; AL023859; CAA19577.1; -
KW Endonuclease.
SQ SEQUENCE 599 AA; 68816 MW; C09ED7219D5895BE CRC64;

```

```

Query Match 29.0%; Score 1011.5; DB 3; Length 599;
Best Local Similarity 36.6%; Pred. No. 2.5e-60;
Matches 241; Conservative 88; Mismatches 149; Indels 181; Gaps 18;

QY 3 SRKSKAAALDTPQSESTFSSTLDSAPSARNLRSGRNILQPSSEKDRDHEKRSGEEL 62
DB 110 ARSSKKKATDSYSDKIDESVASYDST-----HLRSSR-----SKPVNYSSESE- 157
QY 63 AGRMMKDGANGHCLRGKEQEEGVKMAIEGLARMERLQORATKROKKQLEED----- 114
DB 158 -----SEQISKATKKVKQKEEYVEVDK 184
QY 115 GIPVPVWSRF-PTAPYHHKSTNAEREAKPEVLKTHSKDVEREAEIGVDDVVKMEPAAT 173
DB 185 SLANESSDEFEFVPEQLETPISKRRS-----RSSAKNLEKESTMNLDHA----- 232
QY 174 NIEPEDAADAERGAARPPAVNSSLPLPWKGRIGYACLTLYLRNAKPPPIFSRTRCMA 233
DB 233 ----PREMPDCLDK-----PIWGRIGYACLTILRSMKERVFCSTCRIT 275
QY 234 SIYDHRHPLQFDEPEBHLKNNPKDSKEPQDELGHKGFVOELGLANARDIVKMLCWNKYG 293
DB 276 TI-----QRDGLSVKQLGTQNVLDLILKLVENHNEG 307
QY 294 IRELRLSSEMPFASHPVHGYKLAPFASVLAELAGRAVAELGHLRTHHPGQFTOLGSPRK 353
DB 308 IHEMRYSSDLFPFASHAKYGYTL-EFAQSHLEEVGKLANKNHRLTMHPGQVTOIASPRE 366
QY 354 EVVESAIRDLEYHDELLSLLKLPEQONRDVMIHMGQFGDKAATLRFKRNRYARLSOS 413
DB 367 VVDSAIRDLAYHDEILSRMKLNQNLDAVLIHLLGHTFEGKKTLDPRKNYQRLSOS 426
QY 414 CKNRVLNDDVGVTHDILLPVCEELNPMVLDYHHNICFDPHAHLREGTLDISDPKLOE 473
DB 427 VKARLVLEDDVSVSQDILLPLQCELNIPVLDWHHNIIV--PGTLREGSIDLM--PLIP 482
QY 474 RIANTWRKKGKOKMHYSEPCD-GAVTPDRKRHRPRVMTLPPCPDMDLMIEAKDKQEA 532
DB 483 TIRETWTRKGITQKHYSASADPTAISGMKRRASDRVDFPDCDPTMDLMIETAKKEQA 542
QY 533 VFELMRTFKLPGFEKINDMVPYDRDDENRPPVAPKPKKGGKRRKTTDEEAAPEEVD 592
DB 543 VFELCRRYEL-----QNPPCP-----LEIMGPE--- 565
QY 593 TAADDVKDAPEGKPEVEERAMGGPNRYWPLGCEWLKPKKREVKKGKVPVEEVEDE 651
DB 566 --YDQTRDG-----YPPGAEKRLTARKRSRK-----EEVEED 597

```

```

RESULT 3
QYK9P8 PRELIMINARY; PRT; 322 AA.
AC QYK9P8
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE RH2597 PROTEIN.
GN RH2597.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";

```



```
RESULT 10
076891 ID 076891 PRELIMINARY; PRT: 5327 AA.
AC 076891;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE EG:49B4.1 PROTEIN.
GN FUTSCH OR EG:49B4.1 OR CG3064.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Benos P.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL031128; CAA20006.1; -.
DR FlyBase: FBgn0015390; futsch.
SQ SEQUENCE 5327 AA; 575942 MW; FFEER23A118FF38A CRC64;

Query Match 4.5%; Score 155.5; DB 5; Length 5327;
Best Local Similarity 19.4%; Pred. No. 0.28;
Matches 151; Conservative 104; Mismatches 306; Indels 219; Gaps 34;

QY 3 SRKSKAALDPQSESTFS-----STLSSAPSPARNLRGRNLIQPSSE-KORDHE 55
Db 1378 TEKSPLASKETSRPSATGSKVEDTEQTKKSPVSR-----PESEAKDKSP 1426
QY 56 KRSGE-----ELAGRMGKD---ANGH-----CLREGKEQP-----EGVK-- 87
Db 1427 FASGEASRPESVAESVDEACKAESRESIAKTHKDESSLDKAKEQESRRRESIAESIKPE 1486
QY 88 -----MATEGLARERRLORATKQKQLEEDGIPVPSVSRPTAPYHH----- 132
Db 1487 SGIDEKSALAEKASRPESVTDKSPRSRESIAESIAEKSTDEKSGAPSKASRPGSV 1546
QY 133 -KSTNAEREAKEP-----VLKTHSKDVE-REAR--ICVDVVVKMEPATNIIPEDAQD 183
Db 1547 VESVKDETEKSPRSRESIAESAKPTIEFRVSRPESVIDGIKDESA-----KPESRRD 1601
QY 184 A--AERGAARPPAVNSSYLPPLPWKGRGLYACLNLYLNKAK-PPIFSRTCTRMASIV----- 236
Db 1602 SPLASKEASRPESV-----LESVKDEPIKTEKSRRESVAESFK 1640
QY 237 -----DHRHPLQFED--EPEHHLKNKD-----KSKPEQDELGHKFVQEL----- 274
Db 1641 ADSTKDEKSPKTSKDISRPESAVENVMDAPFKETSRPESAVGSKMDSMKSPRSRESVK 1700
QY 275 -GLANARDIVK--MLCWNEYKIGIRFLR-----LSSEMPFFPASHPVHGYK 315
Db 1701 DGAQSRSTSPASVAESAKDGDADDLKELSRPESTTOSKEAGSIKDEKSPLEASE--EASR 1758
QY 316 LAPFASEVLAEAGVAEAELGHLRTHFGQFTQLGSPKPEVVESAIRDELYHDELLSLKL 375
Db 1759 PASVAESVKDAEKSKESRESRESVAESKSPSPKSPASPAVAESIKD-----EAEKS 1810
QY 376 PEQONDAVMIIHMGOGFGDKAATLRFKRNRYARLSQCKNRNLVLENDYDGVTHDLPV 435
Db 1811 KEESRESV-----AEKSPSPKASRPAVAESIKDEAEKSKEE-----SRRESV 1856
QY 436 CEELNPWLDYHHNICFPAHLREGTLDISDPLQERIANTWKRKIGIKQMHYSEPCD 495
Db 1857 AEKSPPL-----SKASRPAVAESIKDEAEKSKESRESRESVAESKSPSPKASRPAS 1909
QY 496 GAVTPRD-----RRKHRRPVMTLPQCPP-----DMDLMTAEADKEQAVFELMRTFKLPG 544
```

```
Db 1910 VAESIKDEAEKSKESRRRESVAESKSPSPKASRPAVAESIKDEAEKSKESRR----- 1964
QY 545 FEKINDMVEFYDRDDENKRPAPPV-----KAPKKKGKGRKRTTD-----EAAAPPEEYDT 593
Db 1965 -ESVAEKSPSPKASRPAVAESIKDEAEKSKESRRRESVAESKSPSPKASRPAV-- 2021
QY 594 AADDVKADEPGKPEPEERAMGPGPNRYWPLGCGEELWKPKKREVKKGKVPVEVEDEGE 653
Db 2022 -AESIKDEAEKSKESRRRESV-----AEKSPSPKASRPAVAESIKDEAE 2067

RESULT 11
Q9JXX5 ID Q9JXX5 PRELIMINARY; PRT: 1763 AA.
AC Q9JXX5;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE HISTONE ACETYLTTRANSFERASE QUERKOPF.
GN MORE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Thomas T., Voss A.K., Chowdhury K., Gruss P.;
RT "Querkopf, a MYST family histone acetyltransferase, is required for normal cerebral cortex development.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF222800; AAF26744.1; -.
DR MGI:1858746; Morf.
DR InterPro: IPR001386; Linker_histone.
DR InterPro: IPR002717; MOZ_SAS.
DR InterPro: IPR001965; PHD.
DR Pfam: PF01853; MOZ_SAS; 1.
DR Pfam: PF00628; PHD; 2.
DR SMART: SM00526; H15; 1.
DR SMART: SM00249; PHD; 2.
DR Transferrase.
KW TRANSFERASE.
SQ SEQUENCE 1763 AA; 196879 MW; 3A570C624B8094B0 CRC64;

Query Match 4.4%; Score 152.5; DB 11; Length 1763;
Best Local Similarity 20.0%; Pred. No. 0.1;
Matches 153; Conservative 86; Mismatches 232; Indels 293; Gaps 40;

QY 80 KQDEGVKMAIEGLARMERRLORATKROKKQLEEDGIPVPS---VVSRRPTAPYHHKSTN 136
Db 319 RPKKGRKLLHKEAAQIKRKYAKPIGRPKNLKQRLSVTSDEGMSAFTGRGSPDTDIK 378
QY 137 AEERAKEFVLKTHSKDVEREAIEIGDDVVKM-EPAATNILEPDAQDAEAERGAARPAV 195
Db 379 ISIKOESADSVLGNKELVTEEDL---DVFOAQELSEKTECESGVE---DCGRYPSV 431
QY 196 -----NSYPLPLPWKGRGLYACL-----NTYLRNAK-----PP----- 223
Db 432 IEFQYKEIQTWYSSPYQPYEARLP-KLYLCEPCLKYMKSKNILLRHSSKCGWGFHPANEI 490
QY 224 -----IFSSRTCTRMASIV-VDHR-----HPLQF-----EPEHHLKN 254
Db 491 YRRKOLSVFVVDGNNSKIYQCNLCLLAKLFLDHKTLTYDVEFPFLFYVLTKNDEKCHLVG 550
QY 255 KPDKSKPEQDELGHKFVQELGLANARDIVKMLCWNEYKIGIRFLRSLSEMFPPFASHPVHGY 314
Db 551 YFSKEK-----LC-QQKYNVSCIMLMPQ-----HQROGF 578
QY 315 KLAPFASEVLAEAGVAEAELGHLRTHFGQFTQLGSPKPEVVE-----SAIRDL 363
Db 579 -----GRFLIDFSYLLSKREG---QAGSEPKPLSDGLRSLYLAWKSVI--L 620
```


QY 364 EY-----HDELLSLKLPQONRDAYMI IHMGQFGDKAATLER-----FKRYARLSQSCKN 416
 DB 621 EYLYRHERHISIKAI-----SRATGCPH-----DIATLQHLHMDRRGRFVIRRE 670
 QY 417 RLVLNDDVGWTVHDLPLVCEELINPMVLDYHHHNCIFCPAHLREGTLDISD-----468
 DB 671 KLILGH-----MEKLNKCSRNE-----LDPESLRTPWMLISNAVVSBEER 711
 QY 469 --PKQIERI--ANTWKR--KGIKQKMYSEPCDGAATPRDRKKURPR-----509
 DB 712 EAEKEARLEMAQSCWEKEEQLLSRVSSROSSAKVQSKNYLHSPERRPVAGERQOLL 771
 QY 510 -----VMTLPP--CPDMDLMEAKDKQAVFELMR 538
 DB 772 ELKSESEEEEEDEDEEESEESIQSPPLTRKPO-----SVSIKRKPFFVVKKR 828
 QY 539 TFKLPQFEKINDMV-----PYDRDENRPA-----PVKAPKK-----572
 DB 829 GRKR--RRINSSVTTTISSETTEVLNPEFNSDEERPMQLEPTCEIPVEEGGRKPVLR 885
 QY 573 -----KGGKRRTTDEAAPEEVDTA-----DDVDADPGPKVEPEEERAMGGPN 620
 DB 886 KAFPHPQKKRQTEEEEGEDNHFKAALCRKDVDQDAQHLKSGKDNPELKC-----939
 QY 621 RVVWPLGCE-----EWLKKPKREVKKG-----KVPEEVEDE 651
 DB 940 QVWPKAGKGLSKWKQSKER--KTGFKLNLNLYTPPETPMEPEQO 981

RESULT 12

Q9NFS3 ID Q9NFS3 PRELIMINARY; PRT; 16215 AA.
 AC Q9NFS3;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE D-TITIN
 GN SL5 OR D-TITIN OR CG1915.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Zhang Y.Q., Broadie K.S.;
 RT "Characterization of Drosophila D-titin gene";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
 CC DOMAIN.
 DR EMBL; AJ271740; CAB93524.1; -;
 DR FlyBase; FBgn003432; sls.
 DR InterPro; IPR002106; AA_TRNA_ligase_II.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR003598; Ig_c2.
 DR InterPro; IPR003600; Ig_like.
 DR InterPro; IPR003006; Ig_MHC.
 DR Pfam; PF00041; fn3; 5.
 DR Pfam; PF00047; Ig; 48.
 DR Pfam; PF00018; SH3; 1.
 DR SMART; SM00408; IgC2; 15.
 DR SMART; SM00410; IG_Like; 34.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS00179; AA-TRNA_LIGASE_II_1; UNKNOWN_1.
 DR PROSITE; PS00002; SH3; 1.
 SQ SEQUENCE 16215 AA; 1841509 MW; 242C8765E00F7603 CRC64;

Query Match

Best Local Similarity 4.18; Score 144; DB 5; Length 16215;
 Matches 140; Conservative 104; Mismatches 266; Indels 232; Gaps 33;

QY 4 RKSAALADTPOSESTFSSTLDSSAPARNLRGRNLL-----OPSEKDRD-----H 54
 DB 13842 KKPPEIKKEEPAEVTV-----STKPVVEEATSSIAVIPQPTTEEAADLKITLIE 13894
 QY 55 EKRSGEELAGRMGKDANGHCLREGKEQEE-----GVKMAIEGLARMERLRQ 102
 DB 13895 EETPPQELVQEI---EBEIEIVEEPKAPEEQPTDFTATKDKSEKKPTVEELPEEQVTIQ- 13949
 QY 103 ATKROKKOLEDGIPIVPSVWSR-----FTAPYHHIKSTNAEEREAKEPEVLKTHSKDVE 155
 DB 13950 --KKKK-----APVPEVVEEPAEFLVKPKTPVQVEETEEAKITKSKPV---KEEAA 13998
 QY 156 REAELGVDDVVKMEPAATNII-----EPEDAADAARGAARPAVNSSYLPLPWKGR 208
 DB 13999 AELKVTITEEPTTEPEVQEIIEEIEEEKPAEVIIEVKESQEAIVEDKEVSLPKK--- 14055
 QY 209 GYACLNTYLNAKPIFFSSRCRMASIVDHRHPIQFDEDEPHHLKNKPD-KSKPEQDEL- 266
 DB 14056 -----KPKAPIV-----EPEAEITLKPVKSEEVQEEAK 14085
 QY 267 ---GHKEFVQELGLANARDIVKMLCWNKEYGIRFLRLSSEMPEPFASHPVHGYKLAPFASE 322
 DB 14086 IVKKPKKIDVAVADELTV-----EFEIKKKKPEPEDIV-----DAAIVLKKPEPVAD 14156
 QY 323 VLAEAGRVAAELGRLTHPGQFTOLGSPRKEVVESAIROLEYHDELLSLKLPQONRD 382
 DB 14120 EVIE-----EFEIKKKKPEPEDIV-----DAAIVLKKPEPVAD 14156
 QY 383 AVMIIHMGQFGDKAATLERKRYARLSQSCKNRLV-LENDVVGWTVHDLPLVCEELNI 441
 DB 14157 EVVAEVLTKPKAKTEVTEEFESVD-VKLPKKKERPEVEIEEEI---IEEAVVIRKKPK 14212
 QY 442 PMV-----LDYHHNICDPAH-LREGTLDISDPKLQERIANTWKRGKJ-----485
 DB 14213 PPEPTVEDLEETEFSLSPKPHITNEGV-----EEAATVLKRPVKPTTLDDEAAAE 14263
 QY 486 -----QKMHYSEPCD-GAVTPRDRRKHPRVMTLPPCPDMDLMEAKDKQAVFELMRT 539
 DB 14264 LSIKQEEYEYEGEDIEEFVVSQQRKPKP-----LQITEDEEAYTVKKLR 14310
 QY 540 FK---LPGEKINDM-----VYDRDENRPA-----PPVKAPKKKKGGKRRKRTD 582
 DB 14311 RKQVDIPEYADVENVTFRARTKTKREDVDQEFNIALDSYAEIEISMSGKVKLKKPKIKTF 14370
 QY 583 EEAAPPEEVDTAADDVKDAPEGPKEVPEERAMGGPNRVVWPLGCEW-----631
 DB 14371 SEADAETKI-----IQDPPDGEPIIEIRDEDTIDEVEEP-----EEYFVEELPPDEV 14423
 QY 632 --LKKPKREVKKGVPEEVEDE 651
 DB 14424 FKLKPKKHPKPAYSVQDEEEQ 14445
 RESULT 13
 ID Q9VEU8 PRELIMINARY; PRT; 828 AA.
 AC Q9VEU8;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
 DE CG14896 PROTEIN.
 GN CG14896.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Abayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottiler P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach G.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong F.N., Zhou X., Zhu S., Zhu H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RA "The genome sequence of *Drosophila melanogaster*."
 RT Science 287:2185-2195(2000).
 RL EMBL; AE003713; AAF53319.1;
 DR FlyBase; FBgn0038433; CG14896.
 SQ SEQUENCE 828 AA; 91444 MW; 96B56F3321CB2DC6 CRC64;

Query Match 4.1%; Score 143.5; DB 5; Length 828;
 Best Local Similarity 19.9%; Pred. No. 0.16;
 Matches 162; Conservative 99; Mismatches 279; Indels 273; Gaps 40;

QY 2 PRRKAAALDTPQSESTFSTLSDSSAPARNLRRSGRNILQPSSEKDR-DHEKRSRG 59
 DB 77 PRRKASAKETNTSDKKEKTDVILAPVKE-----VPSLIRKNSDHSKSG 126
 QY 60 -----ELAGRMCKDANGCHLREGKEQ-----EGVKMAIEGLARM----- 96
 DB 127 POEPAPVATLNISKEQPDNDVNVNSENMAPKPKQERPKRRVRKASEDIVEVNAID 186
 QY 97 -----ERLQRA-----TKRKKOLE-----EDGIPVPSVSRPPTA 128
 DB 187 AINPTALASVOEDHLNQAQEMPKRKRKSAELMLANVEEHAKEDHLERPKRGRKPSL 246
 QY 129 PVHKKSTNAER-----EAKPEVL-----KTHSKOVE- 155
 DB 247 DTDHVSQDVPEKPRRGKTTPADRETQPLGDEKQEPSLPVVEPAKTRRNARKASAEV 306
 QY 156 -----RAETGVDDVVKMEPAANILPEPAQDAEAGAAARPAPVNSVYLPFWKG 206
 DB 307 VSTSGEHLAQIERATEPVTESEPAQSELL-PEEGEENKTRRRGRKPTVDTEGP 362
 QY 207 RLGYACLTNLYRNAP-PFESSRTRMASIVDHRHPLQFDE-----PEHLKNKP 256
 DB 363 -----KTPAEFVPLPSHRRGRKAT-----EDEVLPADLAVPTKURGR- 404
 QY 257 DKSEPODELGHKQVQELGANARDIVKMLCWNEKYGIRFLRLSSEM---FPFASHPVH 312
 DB 405 RASMEPE----HK--DELTASSEVVVELP--TAKTTRGRKPTSDMEATIPKPKPSRR 456

QY 313 GYKLAFASAEVLAEGRAVAALCHRLTHHPQF-----TOLGSPRKEVVESNAIRD 362
 DB 457 GRKASASVDEQPAAKKATARRGRKNEAHEDEERGHIDLQDLPTETASPLVDTSGSPSKA 516
 QY 363 LEYHDELLSLIK-----LPEQONKDAV--MIHMGQFCDKAATLRFKRYARLSQCKN 416
 DB 517 SD-AEELTPRRRGRNLPKRYEEAPDPKPHSG-----LRRARKPAATKSLASK- 565
 QY 417 RLVLNDDVGTVDHLLPVCVELNIPVWLDYHHNICFDPAAHLREGTLDISDPKLOERIA 476
 DB 566 --AEPDFVTPSPVNTNPPVKSEDT-----DNTVLSLEPTTSQR- 603
 QY 477 NTWKRGKIK-QKMHYSPCCGAV-TPDRRKHR-----PRVMT- 512
 DB 604 ----REGNLPKRYNTELDDEMPTPARSRVRNLTAKALELIVDSSPRVTPKPKGKA 659
 QY 513 ----LPPC-----PDMOLMIAKEQAVFELMRTFKLPGEKINDMVPYDRDD----- 558
 DB 660 ANSEEPKAKKPTPEFESTTEAAGPEH---EPIPATKGRGTRRKADDTDLKPDVKTAKKT 716
 QY 559 -----ENRPAPPVKAPKKKKGRKRTTDEEAPEE-----VDTAADVDK 600
 DB 717 VRGAGGKTKVETETEKOPPIKPP---RGARTKTPSEEAPOEEQVKKSAARSRAKGTGA 773
 QY 601 A-PEGKPEVPEERAM-----GGPYNRVYW 624
 DB 774 VEPEEPAEDPQVEASFSSKSTASVRGGRARKVHF 806
 RESULT 14
 Q9H945
 ID Q9H945 PRELIMINARY; PRT; 914 AA.
 AC Q9H945;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE CDNA FLJ13022 FIS, CLONE NT2RP3000753, WEAKLY SIMILAR TO NEUROFILAMENT
 DE TRIPLET H PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Isodai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,
 RA Tanai H., Kimata M., Watanabe M., Hiraoka S., Ishii S., Kawai Y.,
 RA Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K.,
 RA Masuho Y., Kanehori K.;
 RT "NEDO human cDNA sequencing project."
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK023084; BAB14396.1; --
 SQ SEQUENCE 914 AA; 100583 MW; 7E7A9B3A5BAA53C2 CRC64;
 Query Match 4.1%; Score 143.5; DB 4; Length 914;
 Best Local Similarity 19.3%; Pred. No. 0.18;
 Matches 156; Conservative 103; Mismatches 281; Indels 269; Gaps 39;
 QY 11 LDTPOSESSTF-----SSTLSDSSAPARNLRRSGRNILQPSSEKDR-DHEKRSRG-IA 63
 DB 100 VDLPLKESSTFTDENPSETESEAAAGIGLKEGEDGV-KCLSEKDTYDTSIDSLEENLD 158
 QY 64 GRMMGK---DANGCHLREGKEQEGVKMAI-----EGLARMERLQR--ATKROKKOLE 112
 DB 159 KKKKGGKFPESDRCLRSOLSDSSSADRCNRQSSDSACLEIKVPKNPSAKRSKKEGH 218
 QY 113 EDGIPVPSVSRPTAPYHHKSTNAEEREAKPEVLKTHSKDVEREAETGVDDVVKMEPA 172
 DB 219 PGGTTPKGLL-----PDSFHTLETDEKPSVNERPSEKDAEQEGEG---GIITROT 269
 QY 173 TNLIPEDAQDAEAGAAAPPVAVNSYLPPLPWKGLGVACLTNLYRNAPKPPFSSRTRCM 232

```
Db 270 KNNLDEKVGRLGEITFSPSDP-IITAGQPLGERLEIIVQSKMDKNAHIP-----S 320
Qy 233 ASIVDHRHPLQFEDPEH-----HLKKNPKSKPEQDELGHKFKVQELGL 276
Db 321 ESTACKRDEQAKEPGHPIQTHVEAVNEVDNTOOKDDESPCSSLG-----LSS 374
Qy 277 ANARDIVKMLCNWRYGTRFLRSLSEMPFASHPVHGYKLAPF-ASEVLAAGRAVAELG 335
Db 375 SUGSQAARA-----PKSVPRKRLTSTNLR-----HAHSLGLDASKVTS-KEAAQVN 424
Qy 336 HRLTTHPGQFTOLGSPRKE-----VYESAIRDLEYHDELLSLKLPQOONDAVMIHMG 390
Db 425 PIMPRENGA-SESGDPLDEDDVTVVDQPKFMEW-----CAEENQOE--LIANFN 472
Qy 391 GQF-----GDKAATLFRKRYARLSOSCKNRLVLENDVDVGW 427
Db 473 AQYMKVQKGIQLEKEGQTPPRARKSDKLKEIWKSKR-----SRKCRSSLESQK----- 523
Qy 428 TVHDLPLVCEELNIPWLDYHHNIC-----I- 453
Db 524 -----CSPQMLFMTNFKLSNCKWKFLETTETRSILVIVKKLNTRLPGGVPPVKHPLQ 575
Qy 454 -FDPAHLREGTLDISPKLQERI-----ANTWKRKGIGKQMHYS-----EPCDGA-V 498
Db 576 KYAPSSLYPSSLOAE--RLKKHLKFPFGATPAKNWKKQKLWAKFPRENPDQVEPEDGSDV 633
Qy 499 TP-----RDRKRRPRVMTLP-----PCPP 518
Db 634 SPGNSEDSIEEVKEDKDRNSHPANLPTPASTRILRKYSNIRGKLRAQQLIKNEKMECPD 693
Qy 519 DMDLMEADKQOAVFELMRTKLPFGKINDM-----INPLMSPKLALQVDAGFPVKPKSTEGMKGR 738
Db 694 --ALAVESKPSRKSVC----- 568
Qy 569 -----PKKKGGKRRRTTDEAAEPPEVDTAADDVYKDAPE-----GPKVEPEERAM 615
Db 739 KGQVSEILPKAEVQSKRRT---EGSSPDCKNKGPTVKASKEKHADGATKTPAAKRP- 794
Qy 616 GGPYNRVWPLGCEEWLKPKKREVKKGV 644
Db 795 -AADDRSQP-----PKKTSLEKNV 814
```

RESULT 15

Q91255 PRELIMINARY; PRT: 1110 AA.

AC Q91255;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE NF-180.

OS Petromyzon marinus (Sea lamprey).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;

OC Petromyzontiformes; Petromyzontidae; Petromyzon.

OX NCBI_TaxID=7757;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-CENTRAL NERVOUS SYSTEM;

RX MEDLINE=95287814; PubMed=7770000;

RA Jacobs A.J., Kamholz J., Seizer M.E.;

RT "The single lamprey neurofilament subunit (NF-180) lacks

RT multiphosphorylation repeats and is expressed selectively in

RT projection neurons."

RL Brain Res. Mol. Brain Res. 29:43-52(1995).

DR EMBL; U19361; AAA80106.1; -.

DR InterPro; IPR000533; Tropomyosin.

DR InterPro; IPR001664; IF.

DR Pfam; PF00038; filament; 1.

DR PRINTS; PR00194; TROPOMYOSIN.

DR PROSITE; PS00226; IF; UNKNOWN.1.

SO SEQUENCE 1110 AA; 123817 MW; 6558DA73DAF6974C CRC64;

```
Query Match 4.1%; Score 141.5; DB 13; Length 1110;
Best Local Similarity 17.6%; Pred. No. 0.31;
Matches 132; Conservative 102; Mismatches 275; Indels 239; Gaps 24;

Qy 2 PSRKSAAAALDTP-----QSESSTFSTLDSSAPSPARNLRRSGRNLILOPSSSKDRD 53
Db 429 PSAPKAAKVKVSKKPIKVESEPISAQLDLDLEDAOEVEYMEAKAAPVVSAAKDEE 488
Qy 54 HE-----KRSGEELAGRMKGDKANGHCLREGKEQEEGVKMAIEGLARMERRL 100
Db 489 EEEEEEKKEEAEAEAEEDRGRKEGAEEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 548
Qy 101 QRATKROK-----KOLEEDGI--PVPVSVSRPPTAPYHHKHSNABREAA 142
Db 549 EAAEEEEEAGEEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 608
Qy 143 KEPVLKTHSKDVERAEALGVDDVVVKMEPATNIIPEDAODAAERGAARPPAVNSSYLP 202
Db 609 EEEVEAEATKE-EVEAEAEVEE-----EGEAEAEAEAEAEAEAEAEAEAEVTSK----- 651
Qy 203 PWKGRGLVACLNTYLRAKPPIFSRTCRMASIVDHRHPLQFDEPEHHHLKNPKDKSKE- 261
Db 652 -----KANTQAEVEEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 700
Qy 262 --PODELGHKFKVQELGLANARDIVKMLCNWKEYGIRFLRSLSEMFPPASHPVHGYKLAPF 319
Db 701 DAEDEAEAEVEEKE-----EETVK-----SD 721
Qy 320 ASEVLAAGRAVAELGHLRTHPGQFTOLGSPRKEVVESEAIRDLEYHDELLSLKLPQO 379
Db 722 AEEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 769
Qy 380 NRDAVMIHMGQFGDKAATLERKRYARLSQSKNR-----LVLENDVDVGW 427
Db 770 AKEA-----SDDEKPEEEVKESEAPVAPEAKKAPKAPKAPKAPKAPKAPKAPK 814
Qy 428 TVHDLPLVCEELNIPM--VLDYHHHNICFDPALHREGTLDISDPKQLRIANTWKRKGIK 485
Db 815 -----PTSEPEDEPKAEVVE-----KKGAAEAPKPKAPKAPKAPKAPKAPK 854
Qy 486 QKMHYSE-----PCDGAVTPRDRKRHP-----RVMTLPCCPPMDLMI 524
Db 855 KEEPEEPTSEEPKPPAAAKPAKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 914
Qy 525 EAKDKEQAVFELMRTKLPGEKINDMVPYD-----RDENRPPAPVVKAPKKK 573
Db 915 EKEDEDEEE-----EEVEVKPEDAKPVKSKPAPAKAEEDDEPKAPQPPKPKR 962
Qy 574 GGRKRRTTDEAAEP-----EVDTAADDVKDAPGPKPEVPEEERAMGGPYNRVWP 625
Db 963 KPARPKPEEPDKAEPAKKEKHSPEERKPKIKETAKPAKAPAKADKEPA----- 1011
Qy 626 LGCEEWLKPKKREVKKGVPEEVEDEGE 653
Db 1012 -----AEPKKIEVKVKVTKKVVVEIE 1033
```

Search completed: January 15, 2002, 14:06:25
Job time: 964 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:52:22 ; Search time 74.94 Seconds
(without alignments)
377.112 Million cell updates/sec

Title: US-09-724-296-4

Perfect score: 1979
Sequence: 1 DDHAPREMFDCLDKPIPWGR.....KRLTARKRRSRKEVEDEK 371

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1979	100.0	599	S71134	UV-endonuclease -
2	960	48.5	656	S55262	UV-endonuclease -
3	360	18.2	320	S55418	UV-endonuclease ho
4	306.5	15.5	322	E83974	hypothetical prote
5	229.5	11.6	326	C75350	probable UV damage
6	111.5	5.6	517	S45996	H+-transporting AT
7	110.5	5.6	415	B42254	H+-transporting AT
8	110	5.6	465	T44675	H+-transporting AT
9	108.5	5.5	492	A31886	H+-transporting AT
10	108	5.5	500	T14363	probable H+-transp
11	104.5	5.3	472	F69395	H+-transporting AT
12	104	5.3	463	T44310	H+-transporting AT
13	104	5.3	490	S35167	H+-transporting AT
14	103	5.2	494	S18395	H+-transporting AT
15	103	5.2	494	S24387	H+-transporting AT
16	102	5.2	513	C44138	vacuolar H+-transp
17	101.5	5.1	465	C75028	H+-transporting AT
18	101.5	5.1	503	JC4198	adenosinetriphosph
19	101.5	5.1	588	JC7206	phosphoprotein pho
20	101	5.1	491	T34248	hypothetical prote
21	101	5.1	800	B59106	hypothetical prote
22	101	5.1	1145	T33606	hypothetical prote
23	100.5	5.1	511	B44138	hypothetical prote
24	100	5.1	373	C70355	H+-transporting AT
25	100	5.1	513	A33281	8-amino-7-oxonon
26	99.5	5.0	273	E64316	H+-transporting AT
27	99.5	5.0	635	T06619	hypothetical prote
28	98.5	5.0	507	D86207	hypothetical prote
29	97.5	4.9	487	T05679	H+-transporting AT

97.5 4.9 511 2 S32614 H+-transporting AT
97.5 4.9 520 2 G96788 protein T4012.24
32 97 485 2 C86336 hypothetical prote
33 96.5 4.9 465 2 F71213 probable H(+)-tran
34 96.5 4.9 1181 2 T01799 hypothetical prote
35 96 4.9 337 2 S74850 hypothetical prote
36 96 4.9 501 2 S75951 hypothetical prote
37 96 4.9 511 2 S13080 (hypothetical prote
38 94.5 4.8 523 2 D86492 H+-transporting AT
39 94.5 4.8 523 2 D72130 hypothetical prote
40 94 4.7 512 2 T01861 hypothetical prote
41 94 4.7 519 2 D81293 hypothetical prote
42 93.5 4.7 648 2 S16866 probable type II p
43 93.5 4.7 665 2 E70118 gene H19 protein -
44 93 4.7 462 2 G70015 flagellar hook-ass
45 93 4.7 465 2 F69227 conserved hypothet
ATP synthase, subu

ALIGNMENTS

RESULT 1
S71134
UV-endonuclease - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 20-Jun-2000
C:Accession: S71134.139815
R:Takao, M.; Yonemasu, R.; Yamamoto, K.; Yasui, A.
Nucleic Acids Res. 24, 1267-1271, 1996
A:Title: Characterization of a UV endonuclease gene from the fission yeast Schizosacch
A:Reference number: S71134; MUID:96188860
A:Accession: S71134
A:Status: Preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-599 <TAK>
A:Cross-references: EMBL:D78571; NID:g1399000; PIDN:BAAL1415.1; PID:g1399001
A:Experimental source: strain Sp972
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Lucas, M.; Gaillardin, C.
Submitted to the EMBL Data Library, June 1998
A:Reference number: Z21881
A:Accession: T39815
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-599 <LYN>
A:Cross-references: EMBL:AL023859; PIDN:CAA19577.1; GSPDB:GN00067; SPDB:SPBC19C7.09c
A:Experimental source: Strain 972h-; cosmid c19C7
C:Genetics:
A:Gene: UVDE; SPBC19C7.09c
A:Map position: 2

Query Match 100.0%; Score 1979; DB 2; Length 599;
Best Local Similarity 100.0%; Pred. No. 3.4e-146;
Matches 371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DDHAPREMFDCLDKPIPWGRGLGVACLTILRSKRVFCSRTCRITTIORDGLSVKQL 60
DB 229 DDHAPREMFDCLDKPIPWGRGLGVACLTILRSKRVFCSRTCRITTIORDGLSVKQL 288
QY 61 GTQNVLDLILKLVENHNFGIHFMVSSDLFFFAHAKYGYTLEFAQSHLEEVGKLANKYN 120
DB 289 GTQNVLDLILKLVENHNFGIHFMVSSDLFFFAHAKYGYTLEFAQSHLEEVGKLANKYN 348
QY 121 HRLTMHPGQYQTASPREVVDSAIRDLAVHDETLTSMKLNQIKNDAVLIHLGGTFEG 180
DB 349 HRLTMHPGQYQTASPREVVDSAIRDLAVHDETLTSMKLNQIKNDAVLIHLGGTFEG 408
QY 181 KKETLDRFRKNYQRLSDSVKARLVLENDVSVSWQDILLPLCQELNIPVLWDHNNHNVPG 240
DB 409 KKETLDRFRKNYQRLSDSVKARLVLENDVSVSWQDILLPLCQELNIPVLWDHNNHNVPG 468
QY 241 TLREGSLDLMPLPTTRETWTGKTQKHYSADPTAISGMKRAHSDRVDFDFPCDD 300

Db 469 TLREGSLDMLPIPTIRETWRTKGTOKOHYESADPTAISGMKRRRAHSRDVDFPPCDP 528
Qy 301 TMDLMEAKEQAVFELCRVYELQNPCCPLEINGPEVDQTRDGYPPGAEKRLTARKRR 360
Db 529 TMDLMEAKEQAVFELCRVYELQNPCCPLEINGPEVDQTRDGYPPGAEKRLTARKRR 588
Qy 361 SRKEVEDEK 371
Db 589 SRKEVEDEK 599
RESULT 2
S55262
UV-endonuclease - Neurospora crassa
C:Species: Neurospora crassa
C:Date: 10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 20-Jun-2000
C:Accession: S55262
R:Ujima, H.; Takao, M.; Yasuhira, S.; Zhao, J.H.; Ishii, C.; Inoue, H.; Yasui, A.
EMBO J. 14, 2393-2399, 1995
A:Title: A eukaryotic gene encoding an endonuclease that specifically repairs DNA damage
A:Reference number: S55262; MUID:95292980
A:Accession: S55262
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-656 <YAM>
A:Cross-references: EMBL:D11392; NID:g1526560; PID:g927215
C:Genetics:
A:Gene: uvel
Query Match 48.5%; Score 960; DB 2; Length 656;
Best Local Similarity 45.4%; Pred. No. 8.2e-67;
Matches 205; Conservative 54; Mismatches 95; Indels 98; Gaps 10;
Qy 15 PIPWRGLGYACLTILSRMKERVFCSTCRITTI----- 49
Db 201 PLPWKRLGYACLTILSRNKKPPIFSSRTCRMASIVDRHRLQFDEPEHHLKRPDKSK 260
Qy 50 ---QRDGLSVKQLGQNVLDLKLVEWNNHFGIHFMRVSSDLFPFASHAKYGYTL-BFA 105
Db 261 EPODELGHKRFVQELGANARDIVKMLCNEKYGIFRLRSLSEFFAFSHVHYGLAPPA 320
Qy 106 QSHLEEVGKLANKYNHRLTHHPGOYTQIASPREVVVDSAIRDLAYHDEILSRMKLNQIN 165
Db 321 SEVLAEAGRAVAELGHLRTHHPGOYFQLGSPKRVESAIRDLVHDELLSLKLPQON 380
Qy 166 KDAVLIHLGGTTEGKETLDRFRKNYQRLSDSVKARLVLENDVSVQDILLPLCOELN 225
Db 381 RDVAVMIHMGQFGDKAATLERFKRYARLSOSCKRNLVLENDVGVTVHDLPLVCEELN 440
Qy 226 IPLVLDWHHHNIV--PGLREGSLDLM--PLIPTIRETWRTKGTOKOHYESADPTAIS 281
Db 441 IPWLDYHHNICFDPAHLREGTLDISDPKQERIANTRKRGIKOKMYSEPCD-GAVT 499
Qy 282 GMKRRRAHSRDVDFPPCDPTMDLMEAKEQAVFELCRVYEL-----Q 325
Db 500 PRHRRKRRPRVMTLPPCPDMDLMEAKEQAVFELCRVYEL----- 559
Qy 326 NPCC-----LEIMGPE-----YDQTRDG----- 344
Db 560 NRPAPPVAPKPKKKKKRRTTDEEAAPPEVDTAADVDYKDAPEGPKVPEERAMGGPY 619
Qy 345 ---YPPGAEKRLTARKRRSRK-----EEVED 369
Db 620 NRVMYPLGCEEWLKKPKKRVKKGKVPBEVEDE 651
RESULT 3
S55418
UV-endonuclease homolog ywjd - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 15-Oct-1999
C:Accession: S55418; B70060

R:Glaser, P.; Danchin, A.
submitted to the EMBL Data Library, May 1995
A:Description: Cloning and sequencing of the Bacillus subtilis chromosomal region fro
A:Reference number: S55414
A:Accession: S55418
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-320 <GLA>
A:Cross-references: EMBL:249782; NID:g853752; PIDN:CAA9865.1; PID:g853757
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertin, A.M.; Alloni, G.; Azevedo, V.; Ber
C:Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.: Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lechi, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram positive bacterium Bacillus subtili
A:Reference number: A69580; MUID:98044033
A:Accession: B70060
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-320 <KUN>
A:Cross-references: GB:299123; GB:AL009126; NID:g2636240; PIDN:CAB15748.1; PID:el1862
A:Experimental source: strain 168
C:Genetics:
A:Gene: ywjd
Query Match 18.2%; Score 360; DB 2; Length 320;
Best Local Similarity 29.8%; Pred. No. 1.5e-20;
Matches 84; Conservative 61; Mismatches 109; Indels 28; Gaps 8;
Qy 48 TTORDGLSVKQLGQNVLDLKLVEWNNHFGIHFMRVSSDLFPFASH--AKYGYTLEFA 105
Db 34 TERKEALLTVTKANLRNTM---RTLHVIIGHGIPLYRFSISIVPLATHPDMVDFVTFP- 89
Qy 106 QSHLEEVGKLANKYNHRLTHHPGOYTQIASPREVVVDSAIRDLAYHDEILSRMKLNQIN 165
Db 90 QKEPRETEIGELVKTHQLTSFHPNQFTLTSPKESVTNAVTDMAHYRMLEAMGIADR-- 147
Qy 166 KDAVLIHLGGTTEGKETLDRFRKNYQRLSDSVKARLVLENDVSVQDILLPLCOELN 225
Db 148 --SVINIHIGGAYGNKDTATATQAFHONIKOLPQEKERMTLENDKTVTTEETLQVCEQED 205
Qy 226 IPLVLDWHHHNIVPGTULREGSLDLMPLIPTIRETWRTKGTOKOHYESADPTAISGMKR 285
Db 206 VPFVDFPHFVYANP---DDHADLNVALPRMIKTWERTIGLOPKVHLSPPKSEQAI----- 256
Qy 286 RAHSRDV---FDFPPCD-----PTMDLMEAKEQAVFEL 318
Db 257 KSHADYVDANFLLPLLRFRGWTNIDEMIEAKOKDKALLRL 298
RESULT 4
E83974
Hypothetical protein BH2597 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C:Accession: E83974
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A:Reference number: A83650; MUID:20263314
A:Accession: E83974
A:Status: preliminary
A:Molecule type: DNA

A:Accession: A35681
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-34;479-517 <YAM>
A:Accession: B35681
A:Molecule type: protein
A:Residues: 'XV',15-28 <YA2>
R:Nelson, H.; Mandiyan, S.: Nelson, N.
J. Biol. Chem. 264, 5313, 1989
A:Reference number: A40178
A:Contents: erratum
A:Accession: A40178
A:Molecule type: DNA
A:Residues: 478-499, 'TQEAPVRRRTPAKKNL' <NEW>
A:Cross-references: GB:J04450

10

Db 398 SETDRKYLKFADEFEREFIAQRYDEDRSIFETLDLGWELL-----AELPESELKRVKKEY 452
QY 346 ---YPPCAEKR 353
Db 453 ILKYHPKRYKR 463

RESULT 9
A31886
H+-transporting ATPase (EC 3.6.1.35) 57K chain - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 22-Jun-1999
C:Accession: A31886
R:Manolson, M.F.; Ouellette, B.F.F.; Fillion, M.; Poole, R.J.
J. Biol. Chem. 263, 17987-17994, 1998
A:Title: cDNA sequence and homologies of the "57-kDa" nucleotide-binding subunit of the
A:Reference number: A31886; MUID:89053962
A:Accession: A31886
A:Molecule type: mRNA
A:Residues: 1-492 <MAN>
A:Cross-references: GB:J04185; NID:g166626; PIDN:AAC36485.1; PID:g166627
A:Note: it is uncertain whether Met-1, Met-7, or Met-23 is the initiator
C:Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase al
C:Keywords: ATP biosynthesis; hydrogen ion transport; hydrolase; membrane-associated com
F:212-383/Domain: H+-transporting ATP synthase alpha chain homology <ATP>

Query Match 5.5%; Score 108.5; DB 2; Length 492;
Best Local Similarity 20.0%; Pred. No. 1.1;
Matches 69; Conservative 60; Mismatches 123; Indels 93; Gaps 16;
QY 27 LNTILSMKERV-----CSRTCRIITIQDGLSVKQGTQNVLDLIKLVENNH 76
Db 158 MNSIARGQKIPLFSAAGLPHNEIAQICR-----QAGL--VKRL--EKTVDLLE----DH 204

QY 77 ---NFGIHPMVSSDLFPFASHAKYGYTLEFAQSHLEEVG-----KLANK----- 118
Db 205 GEDNFAIVFAAGVNM-----ETAQFFKRFDEENGSMERVTLFLNLANPTIERI 254

QY 119 YNHLRTMHPGQYQTASPREVV-----DSAIRDL-AVHDEILSRMKLNEQLNKDVA 169
Db 255 ITPRIALTAEVLAECGKHLVILDTMSSYADALREVSAAEEVPGRGYPGYMTOLA 314

QY 170 LIHLGGTEGKEKELDRPRKNYQRLSDSVKARLVLENDVSVSDLLPLCQELNPLV 229
Db 315 TIYERAGRIEGRKGSITQT-----PILTMPNDITHTPTPDLTGVTTEGQIYID 362

QY 230 LQWHHNIVPGTLREGSLDMLPLTIRETWTRKGITOKQH-----YSESADPTAISG 282
Db 363 ROLHNQIYV-----PINVLPSLSRLMKSAIGCGMTRKDHSDVSNQLYANYAIGKDYQA 416

QY 283 MKRRAHSDRVDFPPCDPTMDLM-TEAKEQAVPELCRRYELON 326
Db 417 MKAVVGEAL-----SSEDLLEYLFDLKPERKVMOGAYDTRN 454

RESULT 10
T14363
probable H+-transporting ATPase (EC 3.6.1.35) chain B, vacuolar - red alga (Cyanidium ca
N:Alternate names: V-ATPase chain B
C:Species: Cyanidium caldarium
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C:Accession: T14363
R:Ziegler, K.; Hauska, G.; Nelson, N.
Biochim. Biophys. Acta 1230, 202-206, 1995
A:Title: Cyanidium caldarium genes encoding subunits A and B of V-ATPase.
A:Reference number: 216008; MUID:95345110
A:Accession: T14363
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-500 <ZIE>
A:Cross-references: EMBL:U17101; NID:g576660; PIDN:AAA85821.1; PID:g576661
C:Genetics:

A:Introns: 8/3; 73/3; 130/3; 193/1; 294/3; 348/2; 435/2
C:Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase
C:Keywords: hydrolase
F:208-379/Domain: H+-transporting ATP synthase alpha chain homology <ATP>

Query Match 5.5%; Score 108; DB 2; Length 500;
Best Local Similarity 18.7%; Pred. No. 1.2;
Matches 66; Conservative 52; Mismatches 129; Indels 106; Gaps 12;

QY 5 PREMFDCLDKPIPNWRGLGYACLAATILRSMKERV-----CSRTCRIIT-- 48
Db 142 PEEMFE-----TGSIIDVMNSIARGQKIPLFSGAGLPHNEVAAOICQVCLVSTCT 193
QY 49 -IQDGLSVKQGTQNVLDLIKLVENHFGTIHFMRVSSDLFPFASHAKYGYTLEFAQS 107
Db 194 LVKRSKGD-----EEDFAIVFAAGVNM-----ETARFRQ 224

QY 108 HLEEVG-----KLANK-----YNHLRTMHPGQYQTASPREVV-----DSA 144
Db 225 DFEENGAMERVTLFLNLANPTIERIITPRIALTFAEYLAYEKGKHLVILDTMDSAYADA 284

QY 145 IRDL-AVHDEILSRMKLNEQLNKDAVLIHLGGTEGKEKELDRFKNYQRLSDSVKARL 203
Db 285 LREVSAAEEVPGRGYPGYMTDLATIIYERAGRVGRPGSITQL-----PIL 332

QY 204 VLENDVSVSDLLPLCQELNPLVLDWHHNIVPGTLREGSLDMLPLTIRETWTRK 263
Db 333 TMPNDITHTPDLTGVTTEGQIYLDRLQHNQIYV-----PINVLPSLSRLMKSAIGE 386

QY 264 GITOKHYSESADPTAISGKRRKRAHSDRVDFPPCDPTMDLMIEAKEKEQAVF 316
Db 387 GWTRKDHSDVSNQLYAAYAMGKDALMARV-----VGVEALSQEDLLY 429

RESULT 11

F69395
H+-transporting ATP synthase, subunit B (atpB) homolog - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 16-Jul-1999
C:Accession: F69395
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes,
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
A:Reference number: A69250; MUID:98049343
A:Accession: F69395
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-472 <KLE>
A:Cross-references: GB:AE001023; GB:AE000782; NID:g2689346; PIDN:AAB90073.1; PID:g264
C:Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase
F:180-351/Domain: H+-transporting ATP synthase alpha chain homology <ATP>

Query Match 5.3%; Score 104.5; DB 2; Length 472;
Best Local Similarity 20.2%; Pred. No. 2.1;
Matches 51; Conservative 45; Mismatches 93; Indels 63; Gaps 9;

QY 144 AIRDL-AVHDEILSRMKLNEQLNKDAVLIHLGGTEGKEKELDRFKNYQRLSDSVKAR 202
Db 256 ALREISAAREEVPGRGYPGYMTDLATIIYERAGRIGRKGTITQM-----PI 303
QY 203 LVLENDVSVSDLLPLCQELNPLVLDWHHNIVPGTLREGSLDMLPLTIRETWTR 262
Db 304 LTPMGDDITHTPDLTGVTTEGQIIVLSRELHAKGIYV-----PINVLPSLSRLMKEGIG 357
QY 263 KGITOKQH-----YSESAD-----PTAISGKRRKRAHSDRVDFPPCDPTMDLMIEAK 309
Db 358 EGYTRDHPQWMDQYAAVAAEGVDLRLGLVAIVGEEALSERDLF-----LKFA 405

Tue Jan 15 14:12:24 2002

A:Description: Analysis of the gene encoding the 57 KD B-subunit of the vacuolar H⁺-ATPase
 A:Reference number: S25167
 A:Accession: S25167
 A:Molecule type: mRNA
 A:Residues: 1-490 <DOW>
 A:Cross-references: EMBL:X67839; NID:g8809; PIDN:CAA48034.1; PID:g8810
 C:Genetics:
 A:Gene: FlyBase:Vha55
 A:Cross-references: FlyBase:FBgn0005671
 C:Superfamily: H⁺-transporting ATP synthase alpha chain; H⁺-transporting ATP synthase
 C:Keywords: ATP biosynthesis; hydrolase
 F:210-381/Domain: H⁺-transporting ATP synthase alpha chain homology <ATP>

Query Match 5.3%; Score 104; DB 2; Length 490;
 Best Local Similarity 19.7%; Pred. No. 2.4; Indels 86; Gaps 13;
 Matches 61; Conservative 48; Mismatches 114; Indels 86; Gaps 13;

QY 7 EMFDCLDKPI-PW-----RRLGYACLNLTILRSMKRVF-----CSRTCR 45
 DB 128 DFLDIOGQPINWRIYPEEMTGTGSAIDVMSIARGOKIPFSAAGLPHNEIAAQICR 187
 QY 46 ITTQDGLSVKQLGTONVLDLKLIVENHNHFGIHFMRVSSDLFFPASHAKYGYTLEFA 105
 DB 188 -----QAGL--VKLPG-KSVLD-----DHTDNFAIVFAAGVNM-----ETARFF 224
 QY 106 QSHLEEVGKLANK-----YNHRLTMHPGYQTQIASPREVV-----D 142
 DB 225 KQDFENGSMENVCLEFLNLANPTERTIITPLALTAEEFLAYOCEKHLVILTMSSVA 284
 QY 143 SATRD-L-AYHDEILSRMKLEQNLKDAVLIILHGGTFECKKTKETDRFRKNYORLSDSVKA 201
 DB 285 EALREVSAAAREEVPGRKPGGYMYTDLATYERAGRVEGRNGSITQI-----P 332
 QY 202 RLVLNDVSVSVDLLPLCOELNIPVLVDWHHNIIVPGTLREGSLDLMPLIPTIRETWT 261
 DB 333 ILTMPNDIDTHPIPDLTGYITEGQIYVDRQLHNRQIYP-----PVNVLPSLSRLMKSAI 386
 QY 262 RKGITQKQH 270
 DB 387 GEGMTRKDH 395

RESULT 14
 S18395
 H⁺-transporting ATPase (EC 3.6.1.35) chain B, vacuolar - tobacco budworm
 C:Species: Heliothis virescens (tobacco budworm)
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 21-Jan-2000
 C:Accession: S18395
 R:Gill, S.S.; Ross, L.S.
 Arch. Biochem. Biophys. 291, 92-99, 1991
 A:Title: Molecular cloning and characterization of the B subunit of a vacuolar H⁺-ATPase
 A:Reference number: S18395; MUID:92027804
 A:Accession: S18395
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-494 <GIL>
 A:Cross-references: GB:S61797; NID:g237458; PIDN:AAB20098.1; PID:g237459
 C:Superfamily: H⁺-transporting ATP synthase alpha chain; H⁺-transporting ATP synthase
 C:Keywords: hydrolase
 F:214-385/Domain: H⁺-transporting ATP synthase alpha chain homology <ATP>

Query Match 5.2%; Score 103; DB 2; Length 494;
 Best Local Similarity 19.7%; Pred. No. 2.9; Indels 86; Gaps 13;
 Matches 61; Conservative 48; Mismatches 114; Indels 86; Gaps 13;

QY 7 EMFDCLDKPI-PW-----RRLGYACLNLTILRSMKRVF-----CSRTCR 45
 DB 132 DFLDIOGQPINWRIYPEEMTGTGSAIDVMSIARGOKIPFSAAGLPHNEIAAQICR 191
 QY 46 ITTQDGLSVKQLGTONVLDLKLIVENHNHFGIHFMRVSSDLFFPASHAKYGYTLEFA 105

QY 310 EKEQAVFELCRRYELQNPCCPLEI-----MGPEYDQTR-----DGYPYPGAEKRLTAR 357
 DB 406 DEFERRFOOGRYEDRDIEYTLDLGWELLAMLPERELTKVERKFIKYHP-----KYR 458
 QY 358 KRRSRKEVEED 369
 DB 459 KKEAKPEEATEE 470

RESULT 12
 T44310
 H⁺-transporting ATP synthase (EC 3.6.1.34) beta chain [imported] - Thermococcus sp. (str)

C:Species: Thermococcus sp.
 A:Variety: strain KI
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 18-Aug-2000
 C:Accession: T44310
 R:Iida, T.; Kanai, S.; Inatomi, K.; Kamagata, Y.; Maruyama, T.
 Biochim. Biophys. Acta 1329, 12-17, 1997
 A:Title: Alpha- and beta-subunits of a V-type membrane ATPase in a hyperthermophilic sul
 A:Reference number: 222748; MUID:98036007

A:Accession: T44310
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-463 <IID>
 A:Cross-references: EMBL:D88772; NID:g2605626; PIDN:BAA23343.1; PID:g2605628
 A:Experimental source: strain KI
 C:Superfamily: H⁺-transporting ATP synthase alpha chain; H⁺-transporting ATP synthase al
 C:Keywords: ATP biosynthesis; hydrolase; p-loop
 F:181-352/Domain: H⁺-transporting ATP synthase alpha chain homology <ATP>

Query Match 5.3%; Score 104; DB 2; Length 463;
 Best Local Similarity 18.9%; Pred. No. 2.2; Indels 92; Gaps 13;
 Matches 70; Conservative 55; Mismatches 154; Indels 92; Gaps 13;

QY 27 LNTILSRMKRVFCSTCRITITQDGLSVKQLGTONVLDLKLIVENHNHFGIHFMRVS' 86
 DB 139 MNTLVRGQKLPFGSGPLPHNMLAAQIAQAKVLGEE-----EQFAVVF----- 182
 QY 87 SOLFPFASHAKYGYTLE---FAQSHLEEVGKLANK-----YNHRLTMHPG 128
 DB 183 -----AAMGITYEAFNFKKSFEETGAIERAVLFNLADDPATERIITPRMALTV 233
 QY 129 QYQTIASPREVV-----DSAIRDL-AYHDEILSRMKLEQNLKDAVLIILHGGTFE 179
 DB 234 EYLAFDYDMQVLVILTMNTYAEALREISAAREEVPGRKPGGYMYTDLATYERAGRVR 293
 QY 180 GKKTDLRFRKNYQRLSDSVKARLVLENDVSVSVDLLPLCOELNIPVLVDWHHNIIVP 239
 DB 294 GKKGSIQW-----PILTMPDDIDTHPIPDLTGYITEGQIYVLSRELHRKGIYP 341
 QY 240 GTLREGSLDLMPLIPTIRETWTQKQKOH-----YSES---ADPTAISGMKRR 286
 DB 342 -----PIDVLSLSRLMKDGGKGTREDHPQLSQOLYAAAYAGRSRLDLVAVVGEAL 395
 QY 287 AHSDRVDFPPCDPTMDLMIEAKEQAVE-LCRRYELQNPCCPLETMGPEYDQTRDGY 345
 DB 396 SETDRKYLKFADEFERFEFAQRYDEDRSIFETDLGLWELL-----AELPESELKRVKREY 450
 QY 346 ---YPPGAERK 353
 DB 451 ILKYHPKRYKR 461

RESULT 13
 S25167
 H⁺-transporting ATPase (EC 3.6.1.35) chain B, vacuolar - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
 C:Accession: S25167
 R:Dow, J.A.T.; Kaiser, K.
 submitted to the EMBL Data Library, July 1992

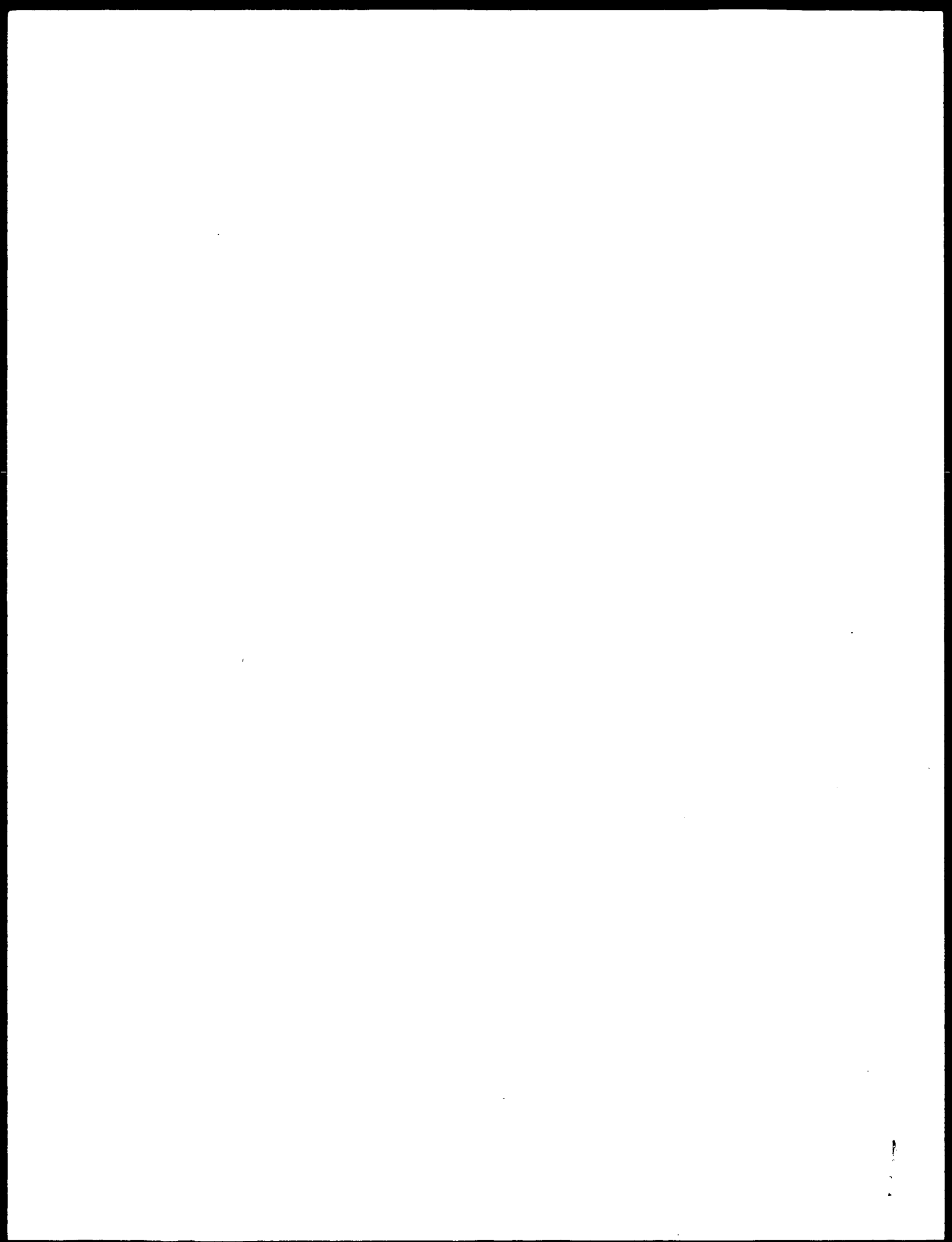
```
Db 192 -----QAGL--VKIPG-KSVLD-----DHEDNFAIVFAAMGVNM-----ETARFF 228
Qy 106 QSHLEEVGKLANK-----YNHRLTMHPGQYQTASPREVV-----D 142
Db 229 KODFEENGSMENVCLFLNLANDPTIERIITPRALTAAEFLAYQCEKHVLVILTMSSYA 288
Qy 143 SAIRDL-AYHDEILSRMKLNQKDAVLIHLGCTEGCKETLDRFRKNYQRLSDSVKA 201
Db 289 EALREVSAAAREEVPGRGFGYMYTDLATYERAGRVEGRNGSITQI-----P 336
Qy 202 RLVLNDDVSNVODLLPLCQELNPLVLDMHHHNIYVPGTLREGSLDMLPIPTIRETWT 261
Db 337 ILTPNDIDITHPIPDLTGYITEGQIYVDRQLHNRQIYP-----PNNVLPSSLRLMKSAI 390
Qy 262 RKGITQKQH 270
Db 391 GEGMTRKDH 399

RESULT 15
S24387
H+-transporting ATPase (EC 3.6.1.35) chain B, vacuolar - tobacco hornworm
C:Species: Manduca sexta (tobacco hornworm)
C:Date: 09-Jun-1994 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
C:Accession: S24387; S19206
R:Novak, F.J.S.; Graef, R.; Waring, R.B.; Wolfersberger, M.G.; Wiczorek, H.; Harvey, W.
Biochim. Biophys. Acta 1132, 67-71, 1992
A:Title: Primary structure of V-ATPase subunit B from Manduca sexta midgut.
A:Reference number: S24387; MUID:92379093
A:Accession: S24387
A:Molecule type: mRNA
A:Residues: 1-494 <NOV>
A:Cross-references: EMBL:X64354; NID:g9713; PIDN:CAA45706.1; PID:g9714
C:Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase al
C:Keywords: ATP biosynthesis; hydrogen ion transport; hydrolase; membrane-associated com
F:214-385/Domain: H+-transporting ATP synthase alpha chain homology <ATP>

Query Match 5.28; Score 103; DB 2; Length 494;
Best Local Similarity 19.7%; Pred. No. 2.9;
Matches 61; Conservative 48; Mismatches 114; Indels 86; Gaps 13;

Qy 7 EMFDCLDKPI-PW-----RRLGYACLNILRSMKERVF-----CSRTCR 45
Db 132 DFLDIQGPINPWSRIYPEMIQTGISAIDVMNSIARGQKIPISAAGLPHNEIAAQICR 191
Qy 46 ITTIQRDGLSVKQLGTQNVLDLIKLVENHNFHFIHFMRVSSDLFPFASHAKYGYTLEFA 105
Db 192 -----QAGL--VKIPG-KSVLD-----DHEDNFAIVFAAMGVNM-----ETARFF 228
Qy 106 QSHLEEVGKLANK-----YNHRLTMHPGQYQTASPREVV-----D 142
Db 229 KODFEENGSMENVCLFLNLANDPTIERIITPRALTAAEFLAYQCEKHVLVILTMSSYA 288
Qy 143 SAIRDL-AYHDEILSRMKLNQKDAVLIHLGCTEGCKETLDRFRKNYQRLSDSVKA 201
Db 289 EALREVSAAAREEVPGRGFGYMYTDLATYERAGRVEGRNGSITQI-----P 336
Qy 202 RLVLNDDVSNVODLLPLCQELNPLVLDMHHHNIYVPGTLREGSLDMLPIPTIRETWT 261
Db 337 ILTPNDIDITHPIPDLTGYITEGQIYVDRQLHNRQIYP-----PNNVLPSSLRLMKSAI 390
Qy 262 RKGITQKQH 270
Db 391 GEGMTRKDH 399
```

Search completed: January 15, 2002, 13:52:23
Job time: 392 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 14:06:13 ; Search time 134.53 Seconds
(without alignments)
403.382 Million cell updates/sec

Title: US-09-724-296-4
Perfect score: 1979
Sequence: 1 DDHAPREMFCDLKPWRG.....KRLTARKRRSRKEEVEDEK 371

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1979	100.0	599	3	Q10988 schizosacch
2	960	48.5	656	3	Q01408 neurospora
3	359	18.1	317	2	Q914f0 bacillus ce
4	359	18.1	325	2	Q914e9 bacillus ce
5	306.5	15.5	322	2	Q9k9p8 bacillus ha
6	229.5	11.6	305	2	Q9s0m9 deinoococcus
7	229.5	11.6	326	2	Q9rte6 deinoococcus
8	111	5.6	496	5	Q9xye8 aedes aegypt
9	109	5.5	492	5	Q44427 culicx quin
10	108	5.5	489	5	Q9u6a8 carcinus ma
11	105	5.3	509	5	Q96066 ascidia syd
12	104	5.3	881	5	Q9gr11 paramacium
13	103	5.2	502	13	Q9w6m4 oncorhynch
14	102.5	5.2	488	10	Q9ase0 oryza sativ
15	101.5	5.1	496	13	Q918a2 gallus gall
16	101.5	5.1	588	3	Q9p982 lentinula e
17	101	5.1	503	13	Q9puk7 anquilla an
18	101	5.1	1145	5	Q9t292 caenorhabdi
19	100.5	5.1	666	2	P96018 staphylococ

20	99.5	5.0	635	10	Q9sz77 arabidopsis
21	99	5.0	511	13	Q9yhdl anquilla an
22	98.5	5.0	507	10	Q9lm14 arabidopsis
23	98	5.0	400	11	Q99ka3 mus musculu
24	97.5	4.9	383	2	Q52231 lactococcus
25	97.5	4.9	383	2	O54680 lactococcus
26	97.5	4.9	487	10	Q9szn1 arabidopsis
27	97.5	4.9	520	10	Q9lqr5 arabidopsis
28	97.5	4.9	793	10	Q9spu5 zea mays (m
29	97	4.9	485	10	Q9lni9 arabidopsis
30	97	4.9	554	2	P94264 borrelia bu
31	97	4.9	634	11	Q99p78 mus musculu
32	97	4.9	822	2	O66172 agrobacteri
33	97	4.9	822	2	Q9r488 agrobacteri
34	97	4.9	1588	11	Q9esk9 mus musculu
35	96.5	4.9	666	2	O54272 staphylococ
36	96.5	4.9	1181	10	O04650 arabidopsis
37	96	4.9	337	2	P73759 synechocyst
38	96	4.9	501	2	O55455 synechocyst
39	95	4.8	633	11	Q99k16 mus musculu
40	95	4.8	634	11	Q99me9 mus musculu
41	95	4.8	780	2	Q916w3 pseudomonas
42	94.5	4.8	523	2	Q9s6p3 chlamydia p
43	94.5	4.8	523	2	Q9z9g2 chlamydia p
44	94	4.7	512	10	O81507 arabidopsis
45	94	4.7	519	2	Q9pmj1 campylobact

ALIGNMENTS

RESULT 1

ID Q10988 PRELIMINARY; PRT; 599 AA.
AC Q10988: P87339;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE UV-ENDONUCLEASE.
GN UVDE OR UVEI+.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP STRAIN=SP972;
RC SEQUENCE FROM N.A.
RX MEDLINE=96188860; PubMed=8614629;
RA Takao M., Yonemasa R., Yamamoto K., Yasui A.;
RT "Characterization of a UV endonuclease gene from the fission yeast
Schizosaccharomyces pombe and its bacterial homolog.";
RL Nucleic Acids Res. 24:1267-1271 (1996).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97175806; PubMed=9023111;
RA Davey S., Nass M.L., Ferrer J.V., Sidik K., Eisenberger A.,
RA Mitchell D.L., Freyer G.A.;
RT "The fission yeast UVDR DNA repair pathway is inducible.";
RL Nucleic Acids Res. 25:1002-1008 (1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Lyne M., Rajandream M.A., Barrell B.G., Lucas M., Gaillardin C.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; D78571; BA011415.1;
DR EMBL; U78487; AAC49664.1;
DR EMBL; AL023859; CAA19577.1;
KW Endonuclease.
SQ SEQUENCE 599 AA; 68816 MW; C09ED7219D5895BE CRC64;

Query Match 100.0%; Score 1979; DB 3; Length 599;

Best Local Similarity 100.0%; Pred. NO. 5.1e-140;		Matches 371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1	DDHAPREMFCDLDRPPIWRGLGYACLTILRSKMERVFCSTCRITTIQDGLGVKQL	60
Db	229	DDHAPREMFCDLDRPPIWRGLGYACLTILRSKMERVFCSTCRITTIQDGLGVKQL	288
Qy	61	GTQNVLDLIKLVENHNFGIHFMRVSSDLPPFASHAKYGYTLEPAQSHLEEVGKLANKYN	120
Db	289	GTQNVLDLIKLVENHNFGIHFMRVSSDLPPFASHAKYGYTLEPAQSHLEEVGKLANKYN	348
Qy	121	HLTMHPGOYTQIASPREVVVDSAIRDLAYHDEILSRMKNEQLNKDAVLIHLGGTFEG	180
Db	349	HLTMHPGOYTQIASPREVVVDSAIRDLAYHDEILSRMKNEQLNKDAVLIHLGGTFEG	408
Qy	181	KKETLDRFKNYQRLSDSVKARLVLENDVSVQDILLPLCOELNIPVLVDWHHNIVPG	240
Db	409	KKETLDRFKNYQRLSDSVKARLVLENDVSVQDILLPLCOELNIPVLVDWHHNIVPG	468
Qy	241	TLREGSLDMLPIPTIRETWTRKGITOKOHYESADPTAISGMKRRASDRVDFPPCDP	300
Db	469	TLREGSLDMLPIPTIRETWTRKGITOKOHYESADPTAISGMKRRASDRVDFPPCDP	528
Qy	301	TMDLMEAKEKEQAVFELCRRYELQNPCCPLEIMGPEYDQTRDGYPPGAEKRLTARKRR	360
Db	529	TMDLMEAKEKEQAVFELCRRYELQNPCCPLEIMGPEYDQTRDGYPPGAEKRLTARKRR	588
Qy	361	SRKEEVEDEK 371	
Db	589	SRKEEVEDEK 599	
RESULT 2			
ID	Q01408	PRELIMINARY; PRT; 656 AA.	
AC	Q01408;		
DT	01-NOV-1996 (TremBLrel. 01, Created)		
DT	01-MAY-1999 (TremBLrel. 10, Last sequence update)		
DT	01-MAY-1999 (TremBLrel. 10, Last annotation update)		
DE	UV-ENONUCLEASE.		
GN	UVEL.		
OS	Neurospora crassa.		
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;		
OC	Sordariales; Sordariaceae; Neurospora.		
OX	NCBI_TaxID=5141;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=74-OR23-1A;		
RX	MEDLINE=95292980; PubMed=7774597;		
RA	Fajima H., Takao M., Yasuhira S., Zhao J.H., Ishii C., Inoue H.,		
RA	Yasui A.;		
RT	"A eukaryotic gene encoding an endonuclease that specifically repairs		
RT	DNA damaged by ultraviolet light."		
RL	EMBO J. 14:2393-2399(1995).		
DR	EMBL; D11392; CAB20113.1; -.		
KW	Endonuclease.		
SQ	SEQUENCE 656 AA; 74413 MW; 4B231E7A8735C86F CRC64;		
Query Match 48.5%; Score 960; DB 3; Length 656;			
Best Local Similarity 45.4%; Pred. NO. 1e-63;			
Matches 205; Conservative 54; Mismatches 95; Indels 98; Gaps 10;			
Qy	15	PIWRGLGYACLTILRSKMERVFCSTCRITTI-----	49
Db	201	PLPWKRLGYACLTILRNAPKPFSSRTRCMASIVDHRHPIQFQDEPPHLLKNKPKDSK	260
Qy	50	---QRDGLSVKQLGTONVLDLIKLVENHNFGIHFMRVSSDLPPFASHAKYGYTL-EPA	105
Db	261	EPQDELGHKFPQVGLGLANARDIVKMLCWNEKYGIRFLRLSEMFPFASHPVHGYKLAPFA	320
Qy	106	QSHLEEVGKLANKYNHLTMHPGOYTQIASPREVVVDSAIRDLAYHDEILSRMKNEQLN	165

Enkaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
Eubrachyura; Portunoidae; Portunidae; Carcinus.
NCBI_TaxID=6759;
[1]
SEQUENCE FROM N.A.
RC
TISSUE=GILL;
RA
Weithrauch D., Lignot J.-H., Towle D.W.;
"Molecular characterization of the B-subunit of a v-type H⁺-ATPase in
gills of the euryhaline shore crab *Carcinus maenas*.";
Am. Zoologist 0:0-0(1999).
[2]

RESULT	11	
ID	096066	
AC	PRELIMINARY;	PRT; 509 AA.
DT	01-MAY-1999	(TREMELrel. 10, Created)
DT	01-MAY-1999	(TREMELrel. 10, Last sequence update)
DT	01-JUN-2001	(TREMELrel. 17, Last annotation update)
DE	VACUOLAR-TYPE H+ ATPASE SUBUNIT B (EC 3.6.1.34).	
GN	VATB.	
OC	Ascidia sydneiensis samea.	
OC	Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Phlebobranchia;	
OC	Ascididae; Ascidia.	
OX	NCBI_TaxID=79730;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=BLOOD;	
RA	Ueki T., Uyama T., Kanamori K., Michibata H.;	
RT	"Isolation of cDNAs encoding A- and B-subunits of the Vacuolar-Type	
RT	ATPase from the ascidian Ascidia sydneiensis samea."	

RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: TO ATPASE ALPHA/BETA CHAINS FAMILY.
 DR EMBL; AB016484; BAA36692.1; -;
 DR InterPro; IPR000194; ATPase_alpha_beta.
 DR InterPro; IPR003255; ATP_synt_B_C.
 DR Pfam; PF000006; ATP_synt_ab; 1.
 DR Pfam; PF01038; ATP_synt_B_C; 1.
 DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
 DR Hydrogen ion transport; Hydrolase.
 KW SEQUENCE 509 AA; 56675 MW; 185BBD824376451B CRC64;
 SQ

Query Match 5.3%; Score 105; DB 5; Length 509;
 Best Local Similarity 17.8%; Pred. No. 6.7;
 Matches 79; Conservative 69; Mismatches 151; Indels 144; Gaps 18;
 QY 5 PREMPDCLDKPIPWGRGL-----GYACNTILRSKMKRVF-----CSRT 43
 DB 132 PEDYLDIOGQPINPKSRIYPEMIQTGISAITMNSIARGQKIPIFSANGLPHEIAAQI 191
 QY 44 CRITTIQDGLSVKOLGTQNVLDLKLVEWNHFGIFHMRVSSDLFPFASHAKYGYTLE 103
 DB 192 CR-----QGGL---VKLPDKVMD-----SHEDNFAIV-----AAGVYME 225
 QY 104 ---FAQSHLEEVGKLANK-----YNHRUTMHPGOYTQIASPREVV----- 141
 DB 226 AARFFKSDFEQHGSDMNVCLFINLANDPTIERIITPRIALTAEFLAYQCEKHVLVILTD 285
 QY 142 ---DSATRLD-AYHDEILSRMKLNEQLNKDAVLIHLGGTFEGKKEFLDRFRKNYQRLS 196
 DB 286 MSSYAEALREVSAREEVPGRGPGCYWYTLATYERAGRVNGRNGSITOI----- 337
 QY 197 DSVKARLVLENDVSWQDLPCLCOELNIPVLVDWHHNIIVPGTLREGSLDMLPLIPTI 256
 DB 338 ---PILTPNDIDTHPIPLDTGYITEGOIVIDRQLHNRQIYP-----PINVLPSSLRL 387
 QY 257 RETWTIRKGIQKQH-----YSESADPTAISGMKRRRAHSRDVDFPPCDPTMDLMI 306
 DB 388 MKSATGEGTRKDHSDVSNQLYANTAIGKQVQAMKAVGVEEAL-----TADMLYLE 439
 QY 307 -----EAKKEQAVF-----ELCRRYELQNPQCPLEIMGPEYDQTRGYPP 348
 DB 440 FLGKPEKTFIAQGPYENSFVDSLNGWELLRIIF-----PKMKLRIPRNVIYYP- 491
 QY 349 GAERKLTARKRSRKEEVEEDK 371
 DB 492 -----RKKVPSAQDKQEDK 506

RESULT 12
 Q9GRIL PRELIMINARY; PRT; 881 AA.
 AC Q9GRIL;
 DT 01-MAR-2001 (TREMREL. 16, Created)
 DT 01-MAR-2001 (TREMREL. 16, Last sequence update)
 DT 01-MAR-2001 (TREMREL. 16, Last annotation update)
 DE ND9 PROTEIN.
 GN ND9.
 OS Parametium tetraurelia.
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
 OC Parametium.
 OX NCBI_TaxID=5888;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Froissard M., Keller A.M., Cohen J.;
 RT "Nd9p, a novel protein with armadillo-like repeats involved in
 RT exocytosis: physiological studies using allelic mutants in
 RT Parametium."
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ293945; CAC12829.1; -;
 SQ SEQUENCE 881 AA; 104104 MW; 8E27791A74CDD4E6 CRC64;

Query Match 5.3%; Score 104; DB 5; Length 881;
 Best Local Similarity 18.5%; Pred. No. 17;
 Matches 75; Conservative 61; Mismatches 126; Indels 144; Gaps 18;
 QY 6 REMFCDCLDKPIPWGRGLGYACUNTILRSKMKRVFCRITTIQDGL-----LESVKQL 60
 DB 44 KEGFDILE-----AKVSQCIDRLIVS-----SQDTYLFQIHFNKGRPKNVPLKNNI 90
 QY 61 GTQNVLDLKLIVE--WNHNFGIHEMKV-----SSDL 89
 DB 91 HTLERDRLKIFQOGWKTYMYKFLIEVRELPTYKGFDRYNYQVTMLKKPQVELMENQEM 150
 QY 90 FPFASHAKYGYTLEFAQSHLEEVGKLANKYNHRLTMHPGOYT-QIASPREVVVDSAI--- 145
 DB 151 YKFNHVLNAGYNIIFLNSY-----QNTKKGLYRNGKNSQFTLOVSETNELEVLEHMHNH 205
 QY 146 ROLAYHDEILSRMKLNEQLNKDAVLIHLGGTFEGKKEFLDRFRKNYQRLSDSVKARLVL 205
 DB 206 ODQIYTAEYVYVINALSEE--EKYWIHSAPYFK-----RKN-----L 240
 QY 206 ENDDVSWQDLPCLCOELNIPVLVDWHHNIIVPGTLREGSLDMLPLIPTIRETWTRKGI 265
 DB 241 YNDLAEWK-----CWQTRAKVVM---KHODEP-----QOQTKKKR 272
 QY 266 TOKQHYSESADPTAISGMKRRRAHSRDVDFPPCDPTMDLMIKAEKEQAVFELCRYLEQ 325
 DB 273 TTKQAIEME--YAVTMRKRYHPYLENF-----VDIVLTF-----LY 308
 QY 326 NPFCPLEIMGPEYDQTRGYYPGAEKRLTARKRSRKEEVEEDK 371
 DB 309 DPCKVKVDNNEADQD-----QOELEEQLOESENEDGEEK 345
 RESULT 13
 Q9W6M4 PRELIMINARY; PRT; 502 AA.
 ID Q9W6M4;
 AC Q9W6M4;
 DT 01-NOV-1999 (TREMREL. 12, Created)
 DT 01-NOV-1999 (TREMREL. 12, Last sequence update)
 DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
 DE V-TYPE ATPASE B SUBUNIT (EC 3.6.1.34).
 OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
 OX NCBI_TaxID=8022;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=GILL;
 RA Perry S.F., Beyers M.L., Johnson D.A.;
 RT "Cloning and molecular characterization of the trout (Oncorhynchus
 RT mykiss) vacuolar H⁺-ATPase B subunit";
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: TO ATPASE ALPHA/BETA CHAINS FAMILY.
 DR EMBL; AF140022; AAD33861.1; -;
 DR InterPro; IPR000194; ATPase_alpha_beta.
 DR InterPro; IPR003255; ATP_synt_B_C.
 DR Pfam; PF000006; ATP_synt_ab; 1.
 DR Pfam; PF01038; ATP_synt_B_C; 1.
 DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
 DR Hydrogen ion transport; Hydrolase.
 KW SEQUENCE 502 AA; 55738 MW; 0FC8CB6972D22BA0 CRC64;
 SQ

Query Match 5.2%; Score 103; DB 13; Length 502;
 Best Local Similarity 19.4%; Pred. No. 9.3;
 Matches 54; Conservative 49; Mismatches 99; Indels 76; Gaps 11;
 QY 27 LNTILRSKMKRVF-----CSRTCRITTIQDGLSVKOLGTQNVLDLKLVEWNH 76
 DB 170 MNSIARGQKIPIFSAGLPHEIAAQICR-----QAGL--VKK--SKDVMDYS-----DD 215
 QY 77 NFGIHEMRVSSDLFPFASHAKYGYTLEFAQSHLEEVGKLANK-----YNH 121

Db 216 NEAIVFAAMGVNN-----ETARFKSDFEENGSMNVCLFLNLANPTIERIITP 265
QY 122 RLTMHPGQVTOIASPREVV-----DSAIRDL-AYHDEILSRMKLNQKDAVLII 172
Db 266 RLALTSAYLAVOCEKHVLVILTDMSSYAEALREVSAREEVPGRGPGYMYTDLATIY 325
QY 173 HLGTFEGKKTLDPRKKNYQRLSDSVKARLVLENDVSWSDVODLLPLCOELNIPVLVDW 232
Db 326 ERAGRVEGRNGSITQI-----PILTMPNDITHTIPDLTGITYTEGOVYVDRQL 373
QY 233 HHNVIPGTLREGSLDMLPLIPTIRETWTKGITOQKH 270
Db 374 HNROIYP-----PINVLPSLSRLMKSATGEGMTRKDH 405

RESULT 14
ID Q9ASE0 PRELIMINARY; PRT; 488 AA.
AC Q9ASE0;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PUTATIVE HA-TRANSPORTING ATP SYNTHASE.
GN P0456F08.19.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartidae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0456F08";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002901; BAB39419.1; - - - - - G9C6821732198EF1 CRC64;
SQ SEQUENCE 488 AA; 54279 MW; 99C6821732198EF1 CRC64;

Query Match 5.2%; Score 102.5; DB 10; Length 488;
Best Local Similarity 19.5%; Pred. No. 9.8;
Matches 73; Conservative 63; Mismatches 138; Indels 101; Gaps 16;
QY 27 LNTILSRMKRVF-----CSRTCRITTTIORGLSVKQL-GTONVLDLKLVEWN 75
Db 154 MNSIARGOKIPLFSAAGLPHNEIAAQICR-----QAGL--VKRLEKSDNILESE---D 202
QY 76 HNFHGHMRVSSDLFPFASHAKYGYTLFAQSHLEEVG-----KLANK-----YN 120
Db 203 ENFAIVFAAMGVNN-----ETAOFKRDPEENGSMERVTFLFNLANDPTIERIIT 252
QY 121 HRLTMHPGQVTOIASPREVV-----DSAIRDL-AYHDEILSRMKLNQKDAVLI 171
Db 253 PRIALTAEYLAYECGKHVLVILTDMSSYADALREVSAREEVPGRGPGYMYTDLATI 312
QY 172 IHLGTFEGKKTLDPRKKNYQRLSDSVKARLVLENDVSWSDVODLLPLCOELNIPVLVD 231
Db 313 VERAGRIEGRKGSITQI-----PILTMPNDITHTIPDLTGITYTEGOVYVDRQ 360
QY 232 WHHNVIPGTLREGSLDMLPLIPTIRETWTKGITOQKH-----YSESADPTAISGMK 284
Db 361 LHNROIYP-----PINVLPSLSRLMKSATGEGMTRRHSVSNQLYANYAIGKDVAMK 414
QY 285 RRAHSDRVDFPPCDPTMDLM-IEAKEKEQAVFELCRRYELQN-----PPCPL 331
Db 415 AVVGEAL-----SSEDLLYLEFLDKFERKFTVQGYADTRNIFOSLDLAWSLRLIFPR 467
QY 332 EIMGPEYDQTRDGY 346
Db 468 ELLHRIPAKTLDQY 482

RESULT 15
ID Q918A2 PRELIMINARY; PRT; 496 AA.
AC Q918A2;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE VACUOLAR H-ATPASE B SUBUNIT OSTEOCLAST ISOZYME (FRAGMENT).
GN VATB.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Andersen P., Lundberg L.G., Stone D.K.;
RT "Chicken vacuolar H-ATPase B subunit isozyme";
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO ATPASE ALPHA/BETA CHAINS FAMILY.
DR EMBL; U61724; AAF73735.1; - - - - -
DR InterPro; IPR000194; ATPase_alpha_beta.
DR InterPro; IPR003255; ATP_synt_B_c.
DR Pfam; PF00006; ATP_synt_ab; 1.
DR Pfam; PF01038; ATP_synt_B_c; 1.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
KW Hydrogen ion transport; Hydrolase.
FT NON_TER 1
SQ SEQUENCE 496 AA; 55049 MW; 1CC5D9FDA68BA7BA CRC64;

Query Match 5.1%; Score 101.5; DB 13; Length 496;
Best Local Similarity 19.2%; Pred. No. 12;
Matches 61; Conservative 55; Mismatches 111; Indels 91; Gaps 14;
QY 2 DHAP-----REMFCDLCKPIWGRGL-----GYAC---LNTILSRMKRVF----- 39
Db 124 DRGPAVLAEDELDIMGQPINPOCRIYPERMIQTGISAIDGMNSIARGOKIPIFSAAGLPH 183
QY 40 ---CSRTCRITTTIORGLSVKQLGTONVLDLKLVEWNHNFHGMVSSDLFPFASHA 96
Db 184 NEIAAQICR-----QAGL--VKK--SKVDYSE-----ENFAIVFAAMGVNN----- 222
QY 97 KYGYTLFAQSHLEEVGKLANK-----YHRLTMHPGQVTOIASPREVV 141
Db 223 ---ETARFKSDFEENGSMNVCLFLNLANPTIERIITPRIALTAEFLAYOCEKHVLV 279
QY 142 -----DSAIRDL-AYHDEILSRMKLNQKDAVLIHLGTFEGKKTLDPRKKNY 192
Db 280 ILTDMSSYAEALREVSAREEVPGRGPGYMYTDLATIVERAGRVEGRNGSITQI----- 335
QY 193 QRLSDSVKARLVLENDVSWSDVODLLPLCOELNIPVLVDWHHNVIPGTLREGSLDMLPL 252
Db 336 -----PILTMPNDITHTIPDLTGITYTEGOVYVDRQLHNROIYP-----PINVLPS 381
QY 253 IPTIRETWTKGITOQKH 270
Db 382 LSRIMKSATGEGMTRKDH 399

Search completed: January 15, 2002, 14:06:15
Job time: 954 sec

0-327384

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:50:09 ; Search time 120.01 Seconds
(without alignments)
357.991 Million cell updates/sec

Title: US-09-724-296-38
Perfect score: 3055
Sequence: 1 MCTTGLSLSLGRCAAPTV.....HKLNKYHDWLWENHEKLSLS 580

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A_Geneseq_1101.*
- 1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
 - 2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
 - 3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
 - 4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
 - 5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
 - 6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
 - 7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
 - 8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
 - 9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
 - 10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
 - 11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
 - 12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
 - 13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
 - 14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
 - 15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
 - 16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
 - 17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
 - 18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
 - 19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
 - 20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
 - 21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
 - 22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3055	100.0	580	AAW74473	Human MED1 endonuc
2	2958.5	96.8	585	AA144504	Human delta228-UV
3	2194.5	71.8	439	AA176348	Human ovarian tumor
4	1230	40.3	257	AAW88701	Secreted protein e
5	194	6.4	219	AAAG02051	Human secreted pro
6	148.5	4.9	326	AAAG16859	Arabidopsis thalia
7	148.5	4.9	327	AAAG16858	Arabidopsis thalia
8	148.5	4.9	352	AAAG16857	Arabidopsis thalia
9	145	4.7	1644	AAW25049	BRCA2 cancer suscep
10	145	4.7	3418	AAW19211	Human breast cance
11	145	4.7	3418	AAW23287	Human breast and o

12	145	4.7	3418	20	AAW04356	Human BRCA2 (Omi3)
13	145	4.7	3418	20	AAW04357	Human BRCA2 (Omi4)
14	145	4.7	3418	20	AAW04358	Human BRCA2 (Omi5)
15	145	4.7	3418	20	AAW04354	Human BRCA2 (Omi1)
16	145	4.7	3418	21	AAW77819	BRCA2 protein sequ
17	144	4.7	2329	18	AAW25038	Partial BRCA2 cance
18	143	4.7	3418	20	AAW04355	Human BRCA2 (Omi2)
19	141.5	4.6	782	11	AAW06991	Polypeptide antige
20	137	4.5	560	21	AAW19182	Lipid metabolism p
21	131.5	4.3	285	20	AAW14200	Mouse DNA demethyl
22	131.5	4.3	291	20	AAW14198	Human DNA demethyl
23	131	4.3	754	19	AAW33811	Tat stimulatory fa
24	130.5	4.3	411	20	AAW14197	Human DNA demethyl
25	130.5	4.3	411	22	AAW99915	Human protein sequ
26	130.5	4.3	411	22	AAW64314	Human protein fi.
27	130.5	4.3	411	22	AAW64844	Heart muscle cell
28	130	4.3	649	20	AAW90005	Expressed antigen
29	130	4.3	867	19	AAW98564	H. pylori GPO 175
30	130	4.3	1743	19	AAW98879	H. pylori GPO 175
31	127.5	4.2	281	20	AAW73829	Human prostate tum
32	127.5	4.2	281	20	AAW48439	Human prostate can
33	126	4.1	263	19	AAW74980	Human secreted pro
34	123	4.0	414	20	AAW14199	Mouse DNA demethyl
35	122.5	4.0	995	22	AAW06223	Peptide #4905 enco
36	122.5	4.0	1202	22	AAW17663	Peptide #4097 enco
37	121.5	4.0	1024	21	AAW42350	Arabidopsis thalia
38	121.5	4.0	1105	21	AAW42349	Arabidopsis thalia
39	121.5	4.0	1189	21	AAW42348	Arabidopsis thalia
40	120.5	3.9	2343	20	AAW80989	Canine factor VIII
41	120.5	3.9	2665	22	AAW14533	Peptide #967 enco
42	120.5	3.9	2665	22	AAW26950	Peptide #987 enco
43	120.5	3.9	2665	22	AAW02259	Peptide #941 enco
44	120.5	3.9	3266	21	AAW42491	Human ORF2255
45	119	3.9	646	21	AAW18188	Plasmodium falcipa

ALIGNMENTS

RESULT 1
AAW74473
ID AAW74473 standard; Protein: 580 AA.
XX
AC AAW74473;
XX
DT 19-MAY-1999 (first entry)
XX
DE Human MED1 endonuclease protein sequence.

XX
DE Endonuclease; MED1: human; methyl-CpG binding endonuclease-1;
KW DNA fidelity; DNA manipulation; cancer; fragile X syndrome; therapy;
KW myotonic dystrophy; Huntington's disease; spinocerebellar ataxia;
KW Kennedy's disease; triplet repeat expansion disorder.

XX
OS Homo sapiens.

XX
PN WO9904626-A1.

XX
PD 04-FEB-1999.

XX
PF 28-JUL-1998; 98W-US15828.

XX
PR 28-JUL-1997; 97US-0053936.

XX
PA (FOX-) FOX CHASE CANCER CENT.

XX
PI Bellacosa A;

XX
DR WPI; 1999-142462/12.

XX
DR N-PSDB; AAX22002#.

XX
PT New nucleic acid encoding human endonuclease MED1 involved in DNA mismatch repair - used for diagnosing susceptibility to cancer and

669 May 28 1999

PT fragile X syndrome, and therapeutically

PS Claim 8; Fig 3; 109pp; English.

XX This sequence is the human MED1 endonuclease of the invention.

CC MED1 (for methyl-CpG binding endonuclease-1) is used to screen for

CC specific modulators (potential therapeutic agents particularly mimetics

CC of MED1) and to study interactions involved in maintaining DNA fidelity,

CC for DNA manipulation and to raise antibodies. Susceptibility or

CC predisposition to cancer (particularly colorectal or endometrial,

CC especially hereditary non-polyposis colorectal cancer), or its prognosis,

CC where caused by alterations in the MED1-encoding gene, are identified by

CC sequence comparison, amplification, detecting altered polypeptide, and

CC restriction fragment mapping, hybridisation (particularly to probes

CC specific for a mutant allele). These same methods can also be used to

CC diagnose fragile X syndrome and other diseases (e.g. myotonic dystrophy,

CC Huntington's disease, spinocerebellar ataxia and Kennedy's disease)

CC associated with triplet repeat expansion. The DNA, or its fragments, are

CC used as probes and primers in the above diagnostic methods, also to

CC isolate homologous sequences, as sources of antisense sequences and for

CC gene transfer, particularly to restore drug sensitivity to drug-resistant

CC cancer cells.

XX Sequence 580 AA;

Query Match 100.0%; Score 3055; DB 20; Length 580;
 Best Local Similarity 100.0%; Pred. No. 7 6e-284;
 Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGTGLESLSGDRGAAPTTSERLVPDPNDLRKEDVAMELERVGEDEQMMIKRSSE 60
 DB 1 mgttleslsldrgaapttserserlvpdpndlrkedvamelervgedeeqmmikrsse 60

QY 61 CNPLLOEPITASQFAGTAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPOGLKFRS 120
 DB 61 cnplllepiasaqfagtagtecrksvpcgwerwvkvrlfgktagrfdvfyfispoglklfrs 120

QY 121 KSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNSNNWNLR 180
 DB 121 ksslanylhnkngetslkpedfdftvlskrgiksykdcsmalstshlqngsnwnlrrt 180

QY 181 RSKCKKDVMPFPSSSSELSQESRGLSNFTSTHLLKKEDEGVDDVNFVRKPKGKVTILKG 240
 DB 181 rskckkdvmpfpssselesqesrglsnftsthlkllkedegvddvnmfrkvrpkgkvtllkg 240

QY 241 IPKTKTKGCRKSCSGFVQSDSKRESVCNKADAEPVAKQSOLDRTVCISDAGACGETL 300
 DB 241 ipiktktkgcrkscsgfvqsdskresvcnkdadepepvakqsoldrtvcisdagacgetl 300

QY 301 SVTSEENLVKKERSLSSGNGFCSEQKTSGLINKFCSAKDSEHNEYEDTFLESEIGT 360
 DB 301 svtseenslvkkerslssgngfcseqktsglinkfcsakdsehneyedtfleseeigt 360

QY 361 KVEVVEREHLHTDILKKGSEMDNNSCPTKXDFTEKIFQEDTTPRQIERRKTSLYFSS 420
 DB 361 kvevverehlhdtlilkgsemdnncsptrkdftekiqfqedtprtqierrktslyfss 420

QY 421 KYNKEALSPPRRKPKKWTTPRPFNLVQETLFDHPDWKLLTATLFTNFTSGKMAIPVLWK 480
 DB 421 kynkealspprrkpkkwtpprpfnlvqetlfdhpdkwlltatlftnftsgkmaipvlwk 480

QY 481 FLEKYPSEAVARTADRWDSVSELLKPLGLYDLRAKTIKVSDEYITLTKWKYPIELHGIGKY 540
 DB 481 flekypseavartadrdvdsellsplglydlraktikvsdeyitltkwkypielhgigky 540

QY 541 GNDSYRIFCVNEWKQVHPDHKLNKYHDWLWENHEKISLS 580
 DB 541 gndsyrifcvnewkvhpdhklntyhdwlwenhekisls 580

RESULT 2
 AAY44504

ID AAY44504 standard; Protein; 565 AA.
 XX
 AC AAY44504;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE Human delta228-UV damage endonuclease.
 XX
 KW delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;
 KW glutathione-S-transferase signal peptide; uvell+ gene product;
 KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
 KW abasic site; aplatium diaduct; mismatched nucleotide pairing;
 KW nucleotide alkylation; skin cancer.

XX Homo sapiens.

OS WO9963828-Al.

PN 16-DEC-1999.

PD 08-JUN-1999; 99WO-US12910.

PF 08-JUN-1998; 98US-0088521.

PR 18-MAY-1999; 99US-0134752.

XX (UYEM-) UNIV EMORY.

XX boetsch PW, Kaur B, Avery AM;

XX WPI; 2000-116417/10.

XX A new truncated ultraviolet damage endonuclease for treatment of skin
 PT cancers -

XX Claim 16; Page 60; 133pp; English.

XX The present sequence is human delta228-UV damage endonuclease.
 CC delta228-UVDE is a 288 amino acid deletion of the N-terminal of the
 CC uvel+ gene product. This is expressed in frame with a GST leader sequence
 CC to generate a fusion protein. This provides stable endonuclease fragments
 CC for cleaving a double-stranded DNA molecule that has a distorted
 CC structure resulting from UV radiation damage, a photoproduct, an abasic
 CC site, mismatched nucleotide pairing, aplatium diaduct, an intercalated
 CC molecule or alkylation of a nucleotide. Uvelp can be used in compositions
 CC for internal or topical application and as a therapeutic agent for skin
 CC cancers.

XX Sequence 565 AA;

Query Match 96.8%; Score 2958.5; DB 21; Length 565;
 Best Local Similarity 97.4%; Pred. No. 1.3e-274;
 Matches 565; Conservative 0; Mismatches 0; Indels 15; Gaps 1;

QY 1 MGTGLESLSGDRGAAPTTSERLVPDPNDLRKEDVAMELERVGEDEQMMIKRSSE 60
 DB 1 mgttleslsldrgaapttserserlvpdpndlrkedvamelervgedeeqmmikrsse 60

QY 61 CNPLLOEPITASQFAGTAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPOGLKFRS 120
 DB 61 cnplllepiasaqfagtagtecrksvpcgwerwvkvrlfgktagrfdvfyfispoglklfrs 120

QY 121 KSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNSNNWNLR 180
 DB 121 ksslanylhnkngetslkpedfdftvlskrgiksykdcsmalstshlqngsnwnlrrt 180

QY 181 RSKCKKDVMPFPSSSSELSQESRGLSNFTSTHLLKKEDEGVDDVNFVRKPKGKVTILKG 240
 DB 181 rskckkdvmpfpssselesqesrglsnftsthlkllkedegvddvnmfrkvrpkgkvtllkg 240

QY 241 IPTKTKTKGCRKSCSGFVQSDSKRESVCNKADAEPVAKQSOLDRTVCISDAGACGETL 300
 DB 241 ipiktktkgcrkscsgfvqsdskresvcnkdadepepvakqsoldrtvcisdagacgetl 300

QY 301 SVTSEENSLVKKERSLSSGNSFCSEOKTSGIINKFCSAKDSHNEKYEDTFLSESEIGT 360
 Db 289 ---seenslvkkerslssgnsfcseoktsgiiinkfcsakdsenhekyedflesseigc 345
 QY 361 KVEVVERKEHLHTDILKRGSEMNNCSPTKDFTGKIFQEDTIPRTQIERKTSLYFSS 420
 Db 346 kvevverkehldilkrsgemnnscptkdfgkifqedtiprtqierktslyfss 405
 QY 421 KYNKEALSPRRKAFKWTTPSPFNLOVETLFHDPWKLIIATIFLNRTSGKMAIPVLWK 480
 Db 406 kynkealspprrkafkwtppspfnlovetlfhdpwkliaiflnrtsgkmaipvlwk 465
 QY 481 FLEKPSAEVARTADWDVSELKPLGLYDLRAKTIVKFSDEYLTWKQKYPHELHGIGKY 540
 Db 466 flekpsaeavartadwvselkplglydlraktivkfsdeyltkwkypielhigiky 525
 QY 541 GNDSYRIFCYNKQVHPDHKLKYNHDLWENHEKLSLS 580
 Db 526 gndsyrfcynkwqvhpdhklkynhldwlenheklsis (565)

RESULT 3

AAW76548 standard; Protein: 439 AA.

AC AAW76548;

DT 10-APR-2000 (first entry)

DE Human ovarian tumor EST fragment encoded protein 44.

KW Expressed sequence tag; EST; human; ovarian tumor; anticancer;

XX gene therapy; treatment.

OS Homo sapiens.

XX DE19817557-A1.

XX 21-OCT-1999.

XX 09-APR-1998; 98DE-1017557.

XX 09-APR-1998; 98DE-1017557.

XX (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

XX WPI; 1999-591920/51.

XX N-PSDB; AAZ77469.

XX New nucleic acid sequences expressed in ovarian, and some other, cancer tissues, and derived polypeptides, for treatment of ovarian cancer and identification of therapeutic agents.

PS Claim 25; Page 261; 310pp; German.

XX This invention describes novel nucleic acid (cDNA) sequences (A) which have anticancer activity and are highly expressed in ovarian tumor tissue (and some also in testis and breast cancer tissue). The products of the invention can be used for gene therapy. (A) are used (i) for recombinant expression of polypeptides (B) and (ii) to isolate complete genes. (B) are used (i) to identify agents suitable for treatment of ovarian cancer; (ii) directly for treating this form of cancer (including expression from gene therapy vectors) and (iii) for generation of specific antibodies. (A) are identified by assembling ESTs (expressed sequence tags) from a particular tissue type before comparison of expression patterns. This allows a significantly longer fragment of the gene to be revealed, so should reduce the number of failures associated with the fact that ESTs from different libraries may represent different parts of the same unknown gene, distorting the estimated frequency of occurrence in a particular tissue. AAY76505-y76638 represent protein

CC fragments encoded by the human ovarian tumor cDNA library derived EST
 CC fragments represented in AAZ77450-277572.

XX Sequence 439 AA;

Query Match 71.8%; Score 2194.5; DB 20; Length 439;
 Best Local Similarity 96.0%; Pred. No. 1.6e-201;
 Matches 427; Conservative 3; Mismatches 6; Indels 9; Gaps 3;

QY 95 KQRLFGKTAGRED-VYFISPOGLKFRSSSLANYLHKNGETSLKPEDFDFTVLSKRGTKS 153
 Db 1 kllf--tsskfplisfsspgglkfrssslanylhngetsikpedfdftvlskrgiks 58
 QY 154 RYKDCSMAALTSHLQNSNNNLRTSKCKKDVFMPPSSSELOESRGLSNFTSTHLL 213
 Db 59 rykdcsmaltshlqnqnsnnlnrtrskckkdvfmppssseiqesrglsnftsthl 118
 QY 214 LKDEGCVDDVNFVRKPKGVYTIKGIPIKTKKGRKSCSGFVQSDSKRESVCNKADA 273
 Db 119 lkdegvdvvnfrkvrpkgvtylikgipiktkgcrkscsgfvqsdskresvcnkada 178
 QY 274 ESEPVAKSQLDRTVCISDAGAGETLSVTSEENSLVKKKERSLSSGNSFCSEQKTSII 333
 Db 179 esepvaqsqldrtvcisdagagetlsvtseenslvkkerslssgnsfcseqktsii 238
 QY 334 NKFCSAKDSHNEKYEDTFLSESEIGTKEVVERKEHLHTDILKRGSEMNNCSPTKDF 393
 Db 239 nkfcsakdsenhekyedtflesseigtkevverkehldilkrsgemnnscptkdf 298
 QY 394 TGEKIFQEDTIPRTQIERKTSLYFSSKYNKCALSPRRKAFKWTTPSPFNLOVETLF 453
 Db 299 t-----editprtqierktslyfsskynkcalsprrkafkwtppspfnlovetlf 352
 QY 454 HDPWKLIIATIFLNRTSGKMAIPVLWKFLKYPSEAEVARTADWDVSELKPLGLYDLRA 513
 Db 353 hdpwkliiatiflnrtsgkmaipvlwkflkypseavartadwvselkplglydlra 412
 QY 514 KTVKFSDEYLTWKQKYPHELHGIG 538
 Db 413 ktvksfdeyltkwkypielhig 437

RESULT 4

AAW88701

ID AAW88701 standard; Protein: 257 AA.

XX AAW88701;

XX 01-MAR-1999 (first entry)

XX Secreted protein encoded by gene 168 clone HCFNFI1.

XX Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

OS Homo sapiens.

XX WO9854963-A2.

XX 10-DEC-1998.

XX 04-JUN-1998; 98WO-US11422.

XX 18-DEC-1997; 97US-0070923.

XX 06-JUN-1997; 97US-0048877.

XX 06-JUN-1997; 97US-0048881.

PR 06-JUN-1997; 97US-0048884.
 PR 06-JUN-1997; 97US-0048893.
 PR 06-JUN-1997; 97US-0048896.
 PR 06-JUN-1997; 97US-0048899.
 PR 06-JUN-1997; 97US-0048915.
 PR 06-JUN-1997; 97US-0048949.
 PR 06-JUN-1997; 97US-0048964.
 PR 06-JUN-1997; 97US-0048972.
 PR 06-JUN-1997; 97US-0049020.
 PR 06-JUN-1997; 97US-0049375.
 PR 05-SEP-1997; 97US-0057628.
 PR 05-SEP-1997; 97US-0057635.
 PR 05-SEP-1997; 97US-0057644.
 PR 05-SEP-1997; 97US-0057647.
 PR 05-SEP-1997; 97US-0057650.
 PR 05-SEP-1997; 97US-0057661.
 PR 05-SEP-1997; 97US-0057667.
 PR 05-SEP-1997; 97US-0057761.
 PR 05-SEP-1997; 97US-0057764.
 PR 05-SEP-1997; 97US-0057770.
 PR 05-SEP-1997; 97US-0057775.
 PR 05-SEP-1997; 97US-0057778.
 PR 06-JUN-1997; 97US-0048875.
 PR 06-JUN-1997; 97US-0048878.
 PR 06-JUN-1997; 97US-0048882.
 PR 06-JUN-1997; 97US-0048885.
 PR 06-JUN-1997; 97US-0048894.
 PR 06-JUN-1997; 97US-0048897.
 PR 06-JUN-1997; 97US-0048900.
 PR 06-JUN-1997; 97US-0048916.
 PR 06-JUN-1997; 97US-0048962.
 PR 06-JUN-1997; 97US-0048970.
 PR 06-JUN-1997; 97US-0048974.
 PR 06-JUN-1997; 97US-0049373.
 PR 05-SEP-1997; 97US-0057584.
 PR 05-SEP-1997; 97US-0057589.
 PR 05-SEP-1997; 97US-0057629.
 PR 05-SEP-1997; 97US-0057642.
 PR 05-SEP-1997; 97US-0057645.
 PR 05-SEP-1997; 97US-0057648.
 PR 05-SEP-1997; 97US-0057651.
 PR 05-SEP-1997; 97US-0057662.
 PR 05-SEP-1997; 97US-0057668.
 PR 05-SEP-1997; 97US-0057762.
 PR 05-SEP-1997; 97US-0057765.
 PR 05-SEP-1997; 97US-0057771.
 PR 05-SEP-1997; 97US-0057776.
 PR 06-JUN-1997; 97US-0048876.
 PR 06-JUN-1997; 97US-0048880.
 PR 06-JUN-1997; 97US-0048883.
 PR 06-JUN-1997; 97US-0048892.
 PR 06-JUN-1997; 97US-0048895.
 PR 06-JUN-1997; 97US-0048898.
 PR 06-JUN-1997; 97US-0048901.
 PR 06-JUN-1997; 97US-0048917.
 PR 06-JUN-1997; 97US-0048963.
 PR 06-JUN-1997; 97US-0048971.
 PR 06-JUN-1997; 97US-0049019.
 PR 06-JUN-1997; 97US-0049374.
 PR 05-SEP-1997; 97US-0057627.
 PR 05-SEP-1997; 97US-0057634.
 PR 05-SEP-1997; 97US-0057643.
 PR 05-SEP-1997; 97US-0057646.
 PR 05-SEP-1997; 97US-0057649.
 PR 05-SEP-1997; 97US-0057654.
 PR 05-SEP-1997; 97US-0057656.
 PR 05-SEP-1997; 97US-0057760.
 PR 05-SEP-1997; 97US-0057763.
 PR 05-SEP-1997; 97US-0057769.
 PR 05-SEP-1997; 97US-0057774.
 PR 05-SEP-1997; 97US-0057777.

(HUMA-) HUMAN GENOME SCI INC.
 XX PA
 XX

PI Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA;
 PI Fan P, Feng P, Ferrite AM, Fischer CL, Florence C;
 PI Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW;
 PI Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM;
 PI Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;
 XX WPI: 1999-059865/05.
 DR N-PSDB; AAV84578.
 DR
 XX
 PT New isolated human genes and the secreted polypeptides they encode -
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 XX
 PS Claim 11; Page 564-565; 772pp; English.
 XX
 CC The invention relates to nucleic acid sequences (AAV84411 to AAV84633)
 CC encoding human secreted proteins (AAW8534 to AAW8756). The secreted
 CC protein gene sequences are deposited with the ATCC under deposit numbers
 CC ATCC 97979, 97974, 97975, 97976, 97977, 209007, 209008, 209009, 209010,
 CC 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host
 CC cells comprising recombinant vectors containing the nucleic acid
 CC sequences are used for the recombinant production of the secreted
 CC proteins. The polynucleotide and amino acid sequences are useful for are
 CC useful for preventing, treating or ameliorating medical conditions e.g.
 CC by protein or gene therapy. Pathological conditions can be also
 CC diagnosed by determining the amount of the new polypeptides in a sample
 CC or by determining the presence of mutations in the new polynucleotides.
 CC Specific uses are described for each of the polynucleotides, based on
 CC which tissues they are most highly expressed in, and include developing
 CC products for the diagnosis or treatment of cancer, neurodegenerative
 CC disorders, developmental abnormalities and foetal deficiencies, blood
 CC disorders, tumours, leukemias, diseases of the immune system, autoimmune
 CC diseases, hepatic and renal disease, lymphomas, inflammation, allergies,
 CC ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,
 CC restenosis, prostate diseases, obesity, disorders involving osteoclasts
 CC such as osteoporosis, arthritis or malignancies, diseases of testes,
 CC lung or thymus, digestive/endocrine disorders, infections and AIDS. The
 CC polypeptides are also useful for identifying their binding partners.
 CC The present sequence represents human secreted protein (see descriptor
 CC line for gene number and clone identification).
 XX
 SQ Sequence 257 AA:

Query Match 40.3%; Score 1230; DB 20; Length 257;
 Best Local Similarity 93.1%; Pred. No. 2.le-109;
 Matches 241; Conservative 3; Mismatches 9; Indels 6; Gaps 1;
 QY 160 MAALTSHLQNSNNWNLRTSKCKKDVFPSPSSSELSOESRGLSNFTSTHLLKDEG 219
 DB 1 maaltshlqnsnnwnlrrtskckkdvfpmpsssselqesrglsnftsthlkdeq 60
 QY 220 VDDVFRKVRKPKGVTLKGIPIKTKKGRKSCSGFVQSDSKRESVCNKADAESEPA 279
 DB 61 vddvfrkvrkpkgvtkilkgipikkkkkgrkscsgfvxsdkresvcnkadaesepva 120
 QY 280 OKSOLDRTVCLSDAGACGTLSTVTSENSLVKKKRSYSSGSNFCSEOKTSCIINKFCSA 339
 DB 121 qksqlartvcisdagacgtclstvtseenslvkkkrsyssgsnfcseoktsglnkfcsa 180
 QY 340 KDEHNEKYEDTFLESEETGKVEVVERKEHLHTDILKRGSEMDNNCSPTKRDKFTGEKIF 399
 DB 181 kdsehnekyedtfleseeigtkevverkehlhtdilkrgsemdnnscsptrkdkft----- 235
 QY 400 QEDTTPRTQIERRKTSLYF 418
 DB 236 -editprntdrkknkpvf 253

RESULT 5
 AAG02051
 ID AAG02051 standard; Protein: 219 AA.
 XX

RESULT 7
AAG16858
ID AAG16858 standard; Protein; 327 AA.
XX
AC AAG16858;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 17662.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
PN EP1033405-A2.
PD
PF 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 18-JUN-1999; 99US-0139817.
PR 21-JUN-1999; 99US-0139899.
PR 22-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.

us-09-724-296-38.rag

Tue Jan-15 14:12:16 2002

PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139452.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 28-JUN-1999; 99US-0140991.
 PR 29-JUN-1999; 99US-0141287.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144684.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 4.9%; Score 148.5; DB 21; Length 352;
 Best Local Similarity 20.7%; Pred No. 1.9e-05;
 Matches 62; Conservative 47; Mismatches 102; Indels 89; Gaps 12;

QY 205 SNFTSTHLLKEDGVDDYNFRKVRPKGVTKILKGIPI-----KTKKCGCRKSCSGF 257

XX Homo sapiens.
 OS WO9722689-A1.
 PN 26-JUN-1997.
 XX 17-DEC-1996; 96WO-US19598.
 XX 29-APR-1996; 96US-0639501.
 PR 18-DEC-1995; 95US-0573779.
 PR 20-DEC-1995; 95US-0575359.
 PR 11-JAN-1996; 96US-0585391.
 XX (HSCR-) HSC RES & DEV LP.
 PA (MYRI-) MYRIAD GENETICS I.
 PA (UYPE-) UNIV PENNSYLVANIA.
 PA (ENDO-) ENDO RECH INC.
 XX Couch F, Kamb A, Rommens JM, Simard J, Tavtigian SV;
 PI Weber BL;
 PI WPI; 1997-341680/31.
 DR N-PSDB; AAT69707.
 XX Human breast cancer susceptibility gene BRCA2 - useful for
 PT diagnosing breast cancer and screening for compounds to treat breast
 PT cancer
 XX Claim 1; Pages 90-106; 189pp; English.
 XX The present sequence is the human breast cancer susceptibility gene
 CC BRCA2 product, which can be used to diagnose breast cancer and
 CC screen for compounds to treat breast cancer. BRCA2 can also be used
 CC in gene therapy to restore wild type BRCA2 gene function to a cell,
 CC which has lost its or has altered (i.e. by virtue of a mutation in
 CC BRCA2) BRCA2 gene function.
 XX Sequence 3418 AA;
 SQ

Query Match 4.7%; Score 145; DB 18; Length 3418;
 Best Local Similarity 20.9%; Pred. No. 0.0017;
 Matches 96; Conservative 67; Mismatches 158; Indels 138; Gaps 23;

Qy 36 KEDVAMELERVGEDEQMIMKRSECNPLLOEPIASQAQFATGTCRKSVPVCGWVVK 95
 Db 972 ksdlsnidkipeknndymnkag-----llgpishsfsgsfirtasneiklsehnikk 1026

Qy 96 QRLFGKTAGRDVYFISPOGLKFRSKSSSLANYLHKNGETSL-KPEDFDFTVLKRGIKSR 154
 Db 1027 skmffk-----dieeqyptsl---acveivntaldnqkklspqsin-tvsahlgssv 1077

Qy 155 YKDCSMAALTSHL--NQNSNNSNNLRTSKCKKDVMPPPSSSELSQESRGLSNFTSTHL 212
 Db 1078 vsdckshltpqmlfskqdfnshnl-----tpsqkaeite-----lstill 1118

Qy 213 LLKEDEGVDDVNFRRKVRK-----PKGVVTILK----- 239
 Db 1119 ----eesgsqfefeqlkpsylqkstfevpenqmtlktseecdadlhvimnapsig 1174

Qy 240 -----GIPKK-----TKGCRKSCSGFVOSDKRESVCNKADAESEPAOKSQL 284
 Db 1175 qvdskskfegteivkrfagllkndcnksasgyilde-----nevgrfgrfysahgtkl 1227

Qy 285 D-RVWCTSDAGACCTGTSVTSEENSLVKKERSLSGSNFCSEQKTSGIINKFCSAKDSE 343
 Db 1228 nvsteaigkavklfsdieniseets-----aevhplissskchd-----svvsmf---kien 1277

Qy 344 HNEKYEDTFLSEEE-----IGTKVEVVERKEHLHTDILKRGSEMDNN---CSP 388
 Db 1278 hndk---tvseknkncqllilqnniemttgtfveei-----tenykrntenednkytaa 1327

Qy 389 TRK---DFTGEKIFQEDTIPRTQIERRKTSLYFSSKYN 423
 Db 1328 srnshniefdgsdskndtv---clhkdetdlldtdqun 1363

RESULT 11
 AAW23287
 ID AAW23287 standard; Protein; 3418 AA.
 AC AAW23287;
 XX 21-FEB-1998 (first entry)
 DE Human breast and ovarian cancer susceptibility protein BRCA2.
 XX BRCA2; breast cancer; ovarian cancer; human;
 KW tumour suppressor protein; gene therapy; receptor.
 XX Homo sapiens.
 XX WO9730108-A1.
 XX 21-AUG-1997.
 XX 19-FEB-1997; 97WO-US03340.
 XX 20-FEB-1996; 96US-0603753.
 XX (UYVA-) UNIV VANDERBILT.
 XX (UNIW) UNIV WASHINGTON.
 XX Claire-King M, Holt JT, Jensen RA, Jetton TL, Page DL;
 PI Robinson-Benton CL, Szabo CI, Thompson ME;
 XX WPI; 1997-434733/40.
 DR N-PSDB; AAT84841.
 XX BRCA1 and BRCA2 tumour suppressor gene products - useful to inhibit
 PT breast and ovarian cancer cell growth and tumorigenesis, or treat
 PT gene linked hereditary or sporadic ovarian or breast cancer
 XX Claim 18; Page 72-88; 148pp; English.
 XX This protein comprises the human BRCA2 tumour suppressor protein
 CC that inhibits the growth of breast and ovarian cancer cells. BRCA2
 CC contains a granin, region, indicating that it is a secreted
 CC protein proteins. BRCA2 protein and DNA sequences encoding BRCA2
 CC (see AAT84841) can be used in the treatment of breast and ovarian
 CC cancers. Since the BRCA2 protein is secreted, it can be used to
 CC identify the BRCA2 receptor and hence to identify BRCA2 protein
 CC mimetics which act on the receptor for use in breast and ovarian
 CC cancer treatment. The BRCA1 gene (see AAT84840) and BRCA2 protein
 CC (see AAW23286) have also been characterised. Methods are claimed for
 CC the isolation of BRCA1 or BRCA2 receptors, and for treating or
 CC preventing breast and (gene-linked hereditary or sporadic) ovarian
 CC cancers using BRCA1 and BRCA2 proteins and genes.
 XX Sequence 3418 AA;
 SQ

Query Match 4.7%; Score 145; DB 18; Length 3418;
 Best Local Similarity 20.9%; Pred. No. 0.0017;
 Matches 96; Conservative 67; Mismatches 158; Indels 138; Gaps 23;

Qy 36 KEDVAMELERVGEDEQMIMKRSECNPLLOEPIASQAQFATGTCRKSVPVCGWVVK 95
 Db 972 ksdlsnidkipeknndymnkag-----llgpishsfsgsfirtasneiklsehnikk 1026

Qy 96 QRLFGKTAGRDVYFISPOGLKFRSKSSSLANYLHKNGETSL-KPEDFDFTVLKRGIKSR 154
 Db 1027 skmffk-----dieeqyptsl---acveivntaldnqkklspqsin-tvsahlgssv 1077

QY 155 YKDCSMAALTSHL--NQSNNSNNWNLRTSKCKKDVFMPPSSSSELSQESRGLSNFTSTHLL 212
 Db 1078 vsdcknshltqpmfkskqdfnsnhnl-----tpsqkaeite-----lstill 1118
 QY 213 LLKEDEGVDDVNRKVRK-----PKGKVTILK----- 239
 Db 1119 -----eesgsqfegtfrkpsylqkatsfeypenqmtlktseecrdadlhvimnapsig 1174
 QY 240 -----GPIPKK-----TKGCRKSCSGFVSDSKRESVCNKADAESEPPVAKQSOL 284
 Db 1175 qvdsskqfegtveikrkfagllkndcnksasgyltde-----nevgrfgrfysahgtkl 1227
 QY 285 D-RTVCISDAGAGETLSVTSENSLVKKERSLSSGSNFCSEQKTSIGTIINKFCSAKDSE 343
 Db 1228 nvstealqkavklfsdieniseets---aevhplslsskchd---svvsmf---kien 1277
 QY 344 HNEKYEDTFLESEE-----IGTKVEVVERKEHLHTDILKRGSEMDNN--CSP 388
 Db 1278 hndk---tvseknnkcqlllqnniemttgtfveei-----tenykrntenednkyytaa 1327
 QY 389 TRK-----DFTGKIFQEDTIPRTQIERRKTSIYFSSKYN 423
 Db 1328 srnshnlefdgsdsskndtv---cjhkdettllftdqn 1363

RESULT 12

AA04356

ID AAY04356 standard; Protein; 3418 AA.

XX

AC AAY04356;

XX

DT 21-JUN-1999 (first entry)

XX

DE Human BRCA2 (omi3) protein.

XX

KW Human; BRCA2; genetic testing; protein therapy; haplotype; detection;
 KW gene therapy; breast cancer; ovarian cancer.

XX

OS Homo sapiens.

XX

PN WO9909164-A1.

XX

PD 25-FEB-1999.

XX

PF 14-AUG-1998; 98WO-US16905.

XX

PR 22-MAY-1998; 98US-0084471.

XX

PR 15-AUG-1997; 97US-0055784.

XX

PR 07-NOV-1997; 97US-0064926.

XX

PR 12-NOV-1997; 97US-0065367.

XX

PR 01-MAY-1998; 98US-0071715.

XX

PA (ONCO-) ONCORMED INC.

XX

PI Eskandari T, Jackson GM, Murphy PD, Olson SJ, Park M;

XX

PI Rabin MB, Schryer B, White MB, Yoshikawa M;

XX

DR WPI; 1999-190163/16.

XX

DR N-PSDB; AAX30257.

XX

New coding sequence haplotypes of the human BRCA2 gene - used to
 develop products for determining susceptibility to, detection and
 treatment of breast or ovarian cancer

XX

Claim 19; Page 112-118; 226pp; English.

XX

The present invention describes genomic DNA which contains a BRCA2 gene
 where the first 12 nucleotides beginning exon 5 are 5'-TCCGTGTTCT-3',
 as in sequence (I) (see AAX03249), where nucleotides numbers
 5782-5790 are GTTGTGTT as in sequence (IV) (see AAX30255), and where
 the last 20 nucleotides encoding exon 15 are 5'-CTGCGTGTTCATAAACAG-3',
 as in sequence (II) (see AAX30251) and the first 20 nucleotides
 beginning exon 16 are 5'-CTGATACGTATGCGGTTTC-3' as in sequence (III)

CC (see AAX30253). Products and methods from the present invention can be
 CC used for identifying mutations in the BRCA2 gene leading to
 CC predisposition or higher susceptibility to breast or ovarian cancer. They
 CC can also be used for detection and gene therapy for breast and ovarian
 CC cancers. They can be used in methods for monitoring disease progression,
 CC for determining patients suited for gene and protein replacement
 CC progression, or for detecting the presence or quantifying the amount of a
 CC tumour growth inhibitor following such therapy. The BRCA2 protein,
 CC polypeptides, their functional equivalents, antibodies, and pns may also
 CC be useful in the study of the characteristics of BRCA2 proteins, such as
 CC structure and function of BRCA2 in oncogenesis or subcellular
 CC localisation of BRCA2 protein in normal and cancerous cells. The present
 CC sequence represents human BRCA2 (omi13).

XX SQ Sequence 3418 AA;

Query Match 4.7%; Score 145; DB 20; Length 3418;
 Best Local Similarity 20.9%; Pred. No. 0.0017;
 Matches 96; Conservative 67; Mismatches 158; Indels 138; Gaps 23;

QY 36 KEDVAMELERVGEDEQMIIKRSSECNPLLEQPIASQFATAGTECRKSVPCGWERVVK 95

Db 972 ksdlslnidkipekundyndkwag-----llgplsnhsfggsfrtsankeikisehnikk 1026

QY 96 QRLFGKTAGRFDYVFTSPQGLKFRSKSSSLANYLHKNGETSL-KPEDFDFTVLKRGIKSR 154

Db 1027 skmfik-----dlsegypstl---acvelvntaldnqklkskpsln-tvsahlqssvv 1077

QY 155 YKDCSMAALTSHL--NQSNNSNNWNLRTSKCKKDVFMPPSSSSELSQESRGLSNFTSTHLL 212

Db 1078 vsdcknshltqpmfkskqdfnsnhnl-----tpsqkaeite-----lstill 1118

QY 213 LLKEDEGVDDVNRKVRK-----PKGKVTILK----- 239

Db 1119 -----eesgsqfegtfrkpsylqkatsfeypenqmtlktseecrdadlhvimnapsig 1174

QY 240 -----GPIPKK-----TKGCRKSCSGFVSDSKRESVCNKADAESEPPVAKQSOL 284

Db 1175 qvdsskqfegtveikrkfagllkndcnksasgyltde-----nevgrfgrfysahgtkl 1227

QY 285 D-RTVCISDAGAGETLSVTSENSLVKKERSLSSGSNFCSEQKTSIGTIINKFCSAKDSE 343

Db 1228 nvstealqkavklfsdieniseets---aevhplslsskchd---svvsmf---kien 1277

QY 344 HNEKYEDTFLESEE-----IGTKVEVVERKEHLHTDILKRGSEMDNN--CSP 388

Db 1278 hndk---tvseknnkcqlllqnniemttgtfveei-----tenykrntenednkyytaa 1327

QY 389 TRK-----DFTGKIFQEDTIPRTQIERRKTSIYFSSKYN 423

Db 1328 srnshnlefdgsdsskndtv---cjhkdettllftdqn 1363

RESULT 13

AA04357

ID AAY04357 standard; Protein; 3418 AA.

XX

AC AAY04357;

XX

DT 21-JUN-1999 (first entry)

XX

DE Human BRCA2 (omi4) protein.

XX

KW Human; BRCA2; genetic testing; protein therapy; haplotype; detection;
 KW gene therapy; breast cancer; ovarian cancer.

XX

OS Homo sapiens.

XX

PN WO9909164-A1.

XX

PD 25-FEB-1999.

XX


```

PF 14-AUG-1998; 98WO-US16905.
XX
PR 22-MAY-1998; 98US-0084471.
PR 15-AUG-1997; 97US-0055784.
PR 07-NOV-1997; 97US-0064926.
PR 12-NOV-1997; 97US-0065367.
PR 01-MAY-1998; 98US-0071715.
XX
PA (ONCO-) ONCORMED INC.
XX
XX Eskandari T, Jackson GM, Murphy PD, Olson SJ, Park M;
PI Rabin MB, Schryer B, White MB, Yoshikawa M;
XX
DR WPI; 1999-190163/16.
DR N-PSDB; AAX30258.
XX
XX New coding sequence haplotypes of the human BRCA2 gene - used to
PT develop products for determining susceptibility to, detection and
PT treatment of breast or ovarian cancer
XX
XX Claim 29; Page 133-140; 226pp; English.
XX
XX The present invention describes genomic DNA which contains a BRCA2 gene
CC where the first 12 nucleotides beginning exon 5 are 5'-TCCTGTTCTTCT-3',
CC as in sequence (I) (see AAX30249), where nucleotides numbers
CC 5782-5790 are GTTGTGTT as in sequence (IV) (see AAX30255), and where
CC the last 20 nucleotides encoding exon 15 are 5'-CTGCGTGTCTCATAAACAG-3'
CC as in sequence (II) (see AAX30251) and the first 20 nucleotides
CC beginning exon 16 are 5'-CTGTATAGTATGGCGTTC-3' as in sequence (III)
CC (see AAX30253). Products and methods from the present invention can be
CC used for identifying mutations in the BRCA2 gene leading to
CC predisposition or higher susceptibility to breast or ovarian cancer. They
CC can also be used for detection and gene therapy for breast and ovarian
CC cancers. They can be used in methods for monitoring disease progression,
CC for determining patients suited for gene and protein replacement
CC progression, or for detecting the presence or quantifying the amount of a
CC tumour growth inhibitor following such therapy. The BRCA2 protein,
CC polypeptides, their functional equivalents, antibodies, and PNs may also
CC be useful in the study of the characteristics of BRCA2 proteins, such as
CC structure and function of BRCA2 in oncogenesis or subcellular
CC localisation of BRCA2 protein in normal and cancerous cells. The present
CC sequence represents human BRCA2 (omi4).
XX
XX Sequence 3418 AA;
SQ

Query Match 4.7%; Score 145; DB 20; Length 3418;
Best Local Similarity 20.9%; Pred. No. 0.0017;
Matches 96; Conservative 67; Mismatches 158; Indels 138; Gaps 23;

QY 36 KEDVAMELERVGEDEQMIKRSSECNPLLOEPIASAOFCATAGTCRKSVPFGWERYVK 95
DB 972 ksdinidkipeknndymnkag-----llgpishnsfggsfrtasnKeiklsehnikk 1026
QY 96 QRLFGTAGTDFVYFISPOGLKRSKSSLLANYLHKNGETSL-KPEDFDFTVLISKRGIKSR 154
DB 1027 skmfk-----dieeqytsl----acveivtialdnqkklspqsin-tvsahlgssvv 1077
QY 155 YKDCSMAALFSLH--NQNSNNWNUNLTRSKCKKDVFMPPSSSELOEERGLSNFTSTHL 212
DB 1078 vsdcknshilpqmlfskqdfnsnhl-----tpsqkaeite-----lslil 1118
QY 213 LLKEDEGVDDVNRKVRK-----PKGKVITLK----- 239
DB 1119 ----eesgsgfctfrkpsylqkstfvepqmtilktseeardadhlvimnapsig 1174
QY 240 -----GIPIKK-----TKGCRKSGFGVQSDSKRESVCNKADAESEPVAKSQL 284
DB 1175 qvdskskfegteivkrkfragllkndcnksasgyltde-----nevfgfgyahgtkl 1227
QY 285 D-RTVCTSDACAGCETLUSVTSEENSLVKKERSLSSGNSFCSEQKTSGLINKFCSAKDSE 343
DB 1228 nvstealqkavklfsdieniseets---aevhplisisskchd-----svvsmf---kien 1277

```

344 HNEKYEDTFLESEE-----ICTKVEVVERKEHLHTDILKRGSEMDNN--CSP 388

1278 hndk---tvseknnkcqllqnniemttgtfveei-----tenykrntenednkyytaa 1327

389 TRK----DFTGEXIFQEDTIPRTQIERRKTSLYFSKYN 423

1328 srushnleldgsdsskndtv---ciliikdetdliiftqhn 1363

RESULT 14

AA04358

ID AAY04358 standard; Protein; 3418 AA.

XX

AC AAY04358;

XX

DT 21-JUN-1999 (first entry)

XX

DE Human BRCA2 (omi5) protein.

XX

KW Human; BRCA2; genetic testing; protein therapy; haplotype; detection;

KW gene therapy; breast cancer; ovarian cancer.

XX

OS Homo sapiens.

XX

PN WO9909164-A1.

XX

PD 25-FEB-1999.

XX

PF 14-AUG-1998; 98WO-US16905.

XX

PR 22-MAY-1998; 98US-0084471.

PR 15-AUG-1997; 97US-0055784.

PR 07-NOV-1997; 97US-0064926.

PR 12-NOV-1997; 97US-0065367.

PR 01-MAY-1998; 98US-0071715.

XX

PA (ONCO-) ONCORMED INC.

XX

PI Eskandari T, Jackson GM, Murphy PD, Olson SJ, Park M;

PI Rabin MB, Schryer B, White MB, Yoshikawa M;

XX

DR WPI; 1999-190163/16.

DR N-PSDB; AAX30259.

XX

PS New coding sequence haplotypes of the human BRCA2 gene - used to

PT develop products for determining susceptibility to, detection and

PT treatment of breast or ovarian cancer

XX

PS Claim 20; Page 155-162; 226pp; English.

XX

CC The present invention describes genomic DNA which contains a BRCA2 gene

CC where the first 12 nucleotides beginning exon 5 are 5'-TCCTGTTCTTCT-3'

CC as in sequence (I) (see AAX30249), where nucleotides numbers

CC 5782-5790 are GTTGTGTT as in sequence (IV) (see AAX30255), and where

CC the last 20 nucleotides encoding exon 15 are 5'-CTGCGTGTCTCATAAACAG-3'

CC as in sequence (II) (see AAX30251) and the first 20 nucleotides

CC beginning exon 16 are 5'-CTGTATAGTATGGCGTTC-3' as in sequence (III)

CC (see AAX30253). Products and methods from the present invention can be

CC used for identifying mutations in the BRCA2 gene leading to

CC predisposition or higher susceptibility to breast or ovarian cancer. They

CC can also be used for detection and gene therapy for breast and ovarian

CC cancers. They can be used in methods for monitoring disease progression,

CC for determining patients suited for gene and protein replacement

CC progression, or for detecting the presence or quantifying the amount of a

CC tumour growth inhibitor following such therapy. The BRCA2 protein,

CC polypeptides, their functional equivalents, antibodies, and PNs may also

CC be useful in the study of the characteristics of BRCA2 proteins, such as

CC structure and function of BRCA2 in oncogenesis or subcellular

CC localisation of BRCA2 protein in normal and cancerous cells. The present

CC sequence represents human BRCA2 (omi5).

XX

CC Sequence 3418 AA;

SQ

[illegible]

Search completed: January 15, 2002, 13:50:15
Job time: 304 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:45:11 ; Search time 120.01 Seconds
(without alignments)
511.063 Million cell updates/sec

Title: US-09-724-296-2

Perfect score: 4349

Sequence: 1 MTKLPILGKWKIKGLVQPTR.....KRLTARKRSRKEVEDEK 828

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_1101.*

1:	/SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2:	/SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3:	/SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4:	/SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5:	/SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6:	/SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7:	/SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8:	/SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9:	/SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10:	/SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11:	/SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12:	/SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13:	/SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14:	/SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15:	/SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16:	/SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17:	/SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18:	/SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19:	/SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20:	/SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21:	/SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22:	/SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4349	100.0	828	21 AAY44498	GST signal peptide
2	3092	71.1	600	21 AAY44500	GST signal peptide
3	1979	45.5	371	21 AAY44499	S. pombe delta228-
4	1259.5	29.0	564	16 AAR84565	Trypanosoma cruzi
5	1259.5	29.0	643	16 AAR84568	Trypanosoma cruzi
6	1237	28.4	229	21 AAY44501	Glutathione-S-tran
7	1221.5	28.1	472	16 AAR84569	Trypanosoma cruzi
8	1220	28.1	379	20 AAY31954	Glutathione-S-tran
9	1217.5	28.0	1140	16 AAR81731	GST-SEP fusion pro
10	1216.5	28.0	310	20 AAY31953	Glutathione-S-tran
11	1216.5	28.0	354	16 AAR84566	Trypanosoma cruzi

12	1212	27.9	824	21 AAY52288	Human HCP/GST fusi
13	1208.5	27.8	856	22 AAB70883	Expression vector
14	1207.5	27.8	362	16 AAR79948	G17 antigen. Chim
15	1206	27.7	426	16 AAR87091	Turkey prolactin/G
16	1205.5	27.7	579	21 AAY58676	GST-Cre recombinas
17	1204.5	27.7	324	22 AAB31613	Amino acid sequenc
18	1203	27.7	271	13 AAR21703	CAP-A-B fusion pro
19	1203	27.7	514	19 AAW47004	Glutathione-S-tran
20	1201.5	27.6	352	16 AAR79947	G15 antigen. Chim
21	1201	27.6	282	20 AAW74096	Glutathione-s-tran
22	1199.5	27.6	232	16 AAR72793	Amino acid sequenc
23	1199.5	27.6	245	22 AAB31612	GST/GI transport r
24	1199.5	27.6	247	20 AAW74114	Glutathione-s-tran
25	1199.5	27.6	247	20 AAW74120	GST/GI transport r
26	1199.5	27.6	248	20 AAW74103	GST/GI transport r
27	1199.5	27.6	248	20 AAW74102	GST/GI transport r
28	1199.5	27.6	248	20 AAW74112	GST/GI transport r
29	1199.5	27.6	248	20 AAW74121	GST/GI transport r
30	1199.5	27.6	248	20 AAW74122	GST/GI transport r
31	1199.5	27.6	248	20 AAW74100	GST/GI transport r
32	1199.5	27.6	248	20 AAW74101	GST/GI transport r
33	1199.5	27.6	248	20 AAW74102	GST/GI transport r
34	1199.5	27.6	249	20 AAW74115	GST/GI transport r
35	1199.5	27.6	257	20 AAW74111	GST/GI transport r
36	1199.5	27.6	257	20 AAW74119	GST/GI transport r
37	1199.5	27.6	257	20 AAW74124	GST/GI transport r
38	1199.5	27.6	257	20 AAW74126	GST/GI transport r
39	1199.5	27.6	257	20 AAW74105	GST/GI transport r
40	1199.5	27.6	257	20 AAW74109	GST/GI transport r
41	1199.5	27.6	258	20 AAW74117	GST/GI transport r
42	1199.5	27.6	258	20 AAW74104	GST/GI transport r
43	1199.5	27.6	259	20 AAW74130	GST/GI transport r
44	1199.5	27.6	259	20 AAW74118	GST/GI transport r
45	1199.5	27.6	259	20 AAW74125	GST/GI transport r

ALIGNMENTS

RESULT 1

ID AAY44498

AC AAY44498 standard; Protein; 828 AA.

AC AAY44498:

XX 27-MAR-2000 (first entry)

XX GST signal peptide and S. pombe UVDE fusion protein.

XX GST signal peptide; ultraviolet damage endonuclease; UVDE; Uvelp;

XX UV irradiation; DNA damage; UV radiation damage; fusion protein;

XX skin cancer; glutathione-S-transferase.

XX Schizosaccharomyces pombe.

XX Synthetic.

XX Key

XX Peptide

XX Protein

XX Misc-difference 11

XX Misc-difference 85

XX Misc-difference 147

XX Misc-difference 147

XX WO9963828-A1.

XX 16-DEC-1999.

XX 08-JUN-1999;

XX 99WO-US12910.

SEQ ID NO: 2
IS AN OBLVIOUS CONST
RUCT

08-JUN-1998: 98US-0088521.
 18-MAY-1999: 99US-0134752.
 (UYEM-) UNIV EMORY.
 Doetsch PW, Kaur B, Avery AM;
 WPI: 2000-116417/10.
 N-PSDB; AAZ29859.
 A new truncated ultraviolet damage endonuclease for treatment of skin cancers -
 Claim 7; Page 51; 133pp; English.
 The present sequence is a fusion protein comprising the GST signal peptide and full length ultraviolet damage endonuclease (UVDE) from *Schizosaccharomyces pombe*. UVDE genes are isolated by introducing a foreign cDNA library into a repair-deficient *E. coli* strain and selecting for complemented cells by UV irradiation of the transformants. This provides stable endonuclease fragments for cleaving a double-stranded DNA molecule that has a distorted structure resulting from UV radiation damage, a photoproduct, an abasic site, mismatched nucleotide pairing, platinum diadduct, an intercalated molecule or alkylation of a nucleotide. Uvde1 can be used in compositions for internal or topical application and as a therapeutic agent for skin cancers.
 Sequence 828 AA:
 Query Match 100.0%; Score 4349; DB 21; Length 828;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1 MTKPLTGLGWKIKGLVQTRILLEYLEEKYEHLIYERDEGDKWRNKKFELGEPNLPYY 60
 1 mtkplilgywklglvqtrilleyleekyeehlyerdegdkwrnkfeliglefpnlpyy 60
 61 IGDVNLKTSQMAIIBYIADKHNMLGGCPKRAETSMLEAGVLDIRYGVSRVAYSQDFETL 120
 61 idgdvnlktsqmaliriyadkhnmlggcpkeraelsmlegavldirygvsvriaysqdfetl 120
 121 KVDLFSKLEPMKMFEDRLCHKTYLNGDRVTHPDVPMLYDALDVLYMDPMLCLAFPKLVC 180
 121 kvdlfslkpekmkfmfedrlchktlylngdrvthpdvpmlydalvlymdpmlclafpklvc 180
 181 FKRIEAIQIDKYLKSSKYIAWPLQGWQATGGGDPHPKSDHLVPRGSMRLRLKRNQI 240
 181 fkrrieaiqidkylksskyiawplqgwqatgggdpkshdhlvprgsmrlrlkrlkriqi 240
 241 SKRIVFTILKQAFGNHPCVPVSVCTITYSVRPHCLPDTLKSLLPMSSKTTLSMLPQVNI 300
 241 skrivftilkqafgnhpcvpvsvctitysvrhclpdtlksllpmsskttlsmlpqvni 300
 301 ANSFSAETPVOLKENETELANISGPHKSTSTSTRKARSSKKATDSVSDKIDESVAS 360
 301 ansfsaetpvolkenetelansgphkstststrkarsskkatdsvasdkidesvas 360
 361 YDSTHLLRRSSRKPVNYSSESSESEEQISKATKKVKQKEEEYVEVDEKSLKNSS 420
 361 ydsthllrrssrskpvnysseseseseeqiskatkvkqkeeeeyvevdekslknss 420
 421 SDFEPVVPQLETPISKRRSSAKNLEKSTWMLDDHAPREMFDCIDKPIWKRGLG 480
 421 sdfepvvpqletpiskrrssaknlekestwmlddhapremfcdkdpipwgrglg 480
 481 YACINTILSRMKERVCSRTCRTITTIQDGLSVKQLGTONVLDLKLVEWNHNFIIHFM 540
 481 yacintilsrmkervcsrtcrtittiqdglsvkqlgtonvldlklvewnhnfgiifm 540
 541 RVSSDLFPFASHAKYGYTLFAGSHLEEVGKLANKYNHRLTMHPGGQYTOIASPREVVVDS 600

Db	241	tlregslm	plptiretwt	rgitqk	hyesad	ptaism	krrahsd	rvtfdp	pcdp	300
QY	758	TMDLMIAE	KEAQVFEL	LCRRYELQ	NPPCPLEI	MGPEYDQ	TRDGYPPG	AERLLTARK	R	817
Db	301	tmdlmiae	kedavfel	crryelgn	ppcpleim	gpeyqdt	rgyppgae	xrltar	xr	360
QY	818	SRKEEVEE	DEEK	828						
Db	361	srkeeevee	dek	371						

RESULT	4	
AAR84565	AAR84565	standard; Protein; 564 AA.
ID	AAR84565	
XX	AC	
XX	AC	
XX	AAR84565;	
XX	09-MAY-1996	(first entry)
DT		
DT		
DE	Trypanosoma cruzi	TCR27 polypeptide, AgI5.
XX		
XX	TCR27;	Chagas disease; repeat unit; diagnosis; blood screening;
KW	recombinant;	fusion protein; glutathione-S-transferase.
KW		
XX	Trypanosoma cruzi.	
XX	OS	
XX	Location/Qualifiers	
Key	329...552	
PH	Region	
FT	/label= repeat_region	
FT	/note= "16 of 69 repeat units of 14 amino acids"	
XX		
XX	W09525797-A1.	
PN		

Query Match 29.0%; Score 1259.5; DB 16; Length 564;
Best Local Similarity 57.1%; Pred. No. 1.2e-86;

Matches	270;	Conservative	55;	Mismatches	103;	Indels	45;	Gaps
QY	5	PILGYWKIGLQVPTRLLEYLEEYEEHLVERDEGDKWRNKKFELGLGFPLNLYPIDG	64					
Db	3	pilgywkiglvqptrllleyleeyleehlyerdegdkwrnkfelfglfplnlypidg	62					
QY	65	VKLTSQMATIRYIADKHNNMLGCPKERAEIFSMLEGAVLDIRYGVSRIRAYSKDFTFLKVD	124					
Db	63	vkltsqmatiriyiadkhnmlgcpkeraeismlegavldirygvsrirayskdfctckvdf	122					
QY	125	LSKLPMLKMFEDRLCHIKTYLNGDHVTHPDMFLYDALDVLVYMDPMLDAPFLKVCFKKR	184					
Db	123	lsklpmlkmfedrlchiktylngdhvthpdmflydaldvlymdpmlcldafpklvcfkkr	182					
QY	185	IFAIPOIDKYLKSSKYIAWPLQGWATGCGDHPKPSDHLVPRG---SMLRLLLKRNIIQISK	242					
Db	183	ieaipqidkylksskyiawplqgwatfggdhpkpsd-lvprgspqlqgaennitnsk	241					
QY	243	RIVFTILKOKAFKGNHPCVPSTCTIT-----YSRFHCLPD---TLKSLLPMSSKT	289					
Db	242	k-emptklrekvvkaekekidaainratkleeernqaykaahkaeeekaktfqrllifesen	300					
QY	290	-TLSMLPQWNIG----ANSFSAETPVDLKKENETELANISGPHKKSTSTSTRKRSRSK	343					
Db	301	inlkktrpdavnrdkknsetakt-d-vekqraaeakavetekgraaeat-kvaaeak	358					
QY	344	KKATDSV----SDKTDESVAVSDSTHLRFSRSRKKPVYNNSSSESEEQISKATKKVK	399					
Db	359	rkaaeakavqtekqraaeatkvaeeakqaaeeakv-----etekqraaeatkva	410					
QY	400	QKEEYEVVEVDEKSLKNSSSDPEFPVPEQIETPISKRRRSRSKAKNLEKE	452					
Db	411	eaekqraaeamkvaeeekkaa-----eatkvaeeekqkaaeatkvaeeak	455					

RESULT	5	
AA84568	AA84568 standard; Protein: 643 AA.	
AC	AA84568;	
DT	09-MAY-1996 (first entry)	
DE	Trypanosoma cruzi TCR27 polypeptide, Ag15.	
FW	TCR27; Chagas disease; repeat unit; diagnosis: blood screening;	
KW	recombinant; fusion protein; glutathione-S-transferase.	
OS	Trypanosoma cruzi.	
FT	Key	Location/Qualifiers
FT	Region	329..552
FT		/label= repeat_region
FT		/note= "16 of 69 repeat units of 14 amino acids"
PN	WO9525797-A1.	
XX	28-SEP-1995.	
XX	20-MAR-1995; 95WO-US03191.	
XX	24-MAR-1994; 94US-0216894.	
XX	(KIRC/) KIRCHHOFF L V.	
XX	(OTSU/) OTSU K.	
XX	Kirchhoff LV, Otsu K;	
XX	WPI: 1995-344618/44.	
XX	N-PSDB: AAT05332.	
XX	New polypeptide(s) from the TCR27 protein of Trypanosoma cruzi	
XX	immunossay reagent for specific diagnosis of Chagas disease, alt	

AC AAR84569;
 XX
 DT 09-MAY-1996 (first entry)
 XX
 DE Trypanosoma cruzi TCR27 polypeptide, Ag8.
 XX
 KW TCR27; Chagas disease; repeat unit; diagnosis; blood screening;
 KW recombinant; fusion protein; glutathione-S-transferase.
 XX
 OS Trypanosoma cruzi.
 XX
 FH Key Location/Qualifiers
 FT Region 241..450
 FT /label= repeat_region
 FT /note= "15 of 69 repeat units of 14 amino acids"
 XX
 PN W09525797-A1.
 XX
 PD 28-SEP-1995.
 XX
 PF 20-MAR-1995; 95WO-US03191.
 XX
 PR 24-MAR-1994; 94US-0216894.
 XX
 PA (KIRC/) KIRCHHOFF L V.
 PA (OTSU/) OTSU K.
 XX
 PI Kirchhoff LV, Otsu K;
 XX
 DR WPI: 1995-344618/44.
 DR N-PSDB; AAT05333.
 XX
 XX New polypeptide(s) from the TCR27 protein of Trypanosoma cruzi - as
 PT immunoassay reagent for specific diagnosis of Chagas disease, also
 PT related nucleic acid and transformed cells
 XX
 PS Disclosure; Page 44-45; 68pp; English.
 XX
 CC AAR84565-R84569 are polypeptides of the TCR27 protein of T. cruzi
 CC The proteins are all fusion products with glutathione-S-transferase
 CC (GST) and some contain a linker sequence. The TCR27 protein comprises
 CC a 95 amino acid (aa) N-terminal region; 69 repeats of a highly
 CC conserved 14 aa sequence and a 68 aa C-terminal region. This sequence
 CC encodes the GST sequence, the Ag8 polypeptide contg. 15 of the 69
 CC repeat units and also contains a linker sequence. The TCR27
 CC polypeptides of the invention are useful for the diagnosis of Chagas
 CC disease (American trypanosomiasis), they are capable of detecting
 CC anti-T. cruzi antibodies; or for blood screening. The TCR27 protein
 CC has epitopes to which most T. cruzi infected individuals have
 CC antibodies. The TCR27 polypeptides will not react with serum from
 CC patients with leishmaniasis, schistosomiasis, or autoimmune disease
 CC and are hence less likely to cause false positives in diagnosis.
 XX
 SQ Sequence 472 AA;

Query Match 28.1%; Score 1221.5; DB 16; Length 472;
 Best Local Similarity 56.7%; Pred. No. 6.6e-84;
 Matches 255; Conservative 54; Mismatches 110; Indels 31; Gaps 9;

QY 5 PILGWYKIGLVQPTLLLEYLEEKYEHLHYERDEGDKWRNKKFELGLFPPNLYIDGD 64
 DB 3 PILGWYKIGLVQPTLLLEYLEEKYEHLHYERDEGDKWRNKKFELGLFPPNLYIDGD 62
 QY 65 VKLTQSMATIRYTDKHNMLGGCGPKRAEISMLEGAVLDIRYGVSRVAYSKQFETLKVD 124
 DB 63 VKLTQSMATIRYTDKHNMLGGCGPKRAEISMLEGAVLDIRYGVSRVAYSKQFETLKVD 122
 QY 125 LSKLPEMLKMFEDRLCHTKYLYNGDHYTHPDFMLYDALDVLVMDPNCMLDAFPLKVCFFKR 184
 DB 123 LSKLPEMLKMFEDRLCHTKYLYNGDHYTHPDFMLYDALDVLVMDPNCMLDAFPLKVCFFKR 182
 QY 185 IEATPQIDKYLKSSKYTAWPLOGMQATFGGDDHPKSDHLVPRGSMRLRLKRNLIQISKRT 244

DB 183 IEATPQIDKYLKSSKYTAWPLOGMQATFGGDDHPKSDHLIEIGR-----IPPGCNSTKV 237
 QY 245 VFTILKOKAFKGNHPCVPSVCTITYSRFHCLPDTLKSLLPMSKTKTSLMLPQVNTIGANSF 304
 DB 238 a-eaekqkaaeat-----kvaeeakqraaeatk--vaeaekqkaaeatkvaeeakqr 286
 QY 305 SATETP--VDLKKEKETELANISGPHKKSSTSTRKPARSKKKATDSVSDKIDSVASYD 362
 DB 287 aaeatkvaeeakqkaaeatkvaagdekqkaaeat-kvaeeakqkaaeat--kvae--aekq 341
 QY 363 SSTHLRRSSRSKPKPVNYNSSESESESEQISKATKKVKQKEEYVEEVDEKSLKNESSD 422
 DB 342 kaeeatkvaeeakqkaaeatkvaeeakqkaaeatkvaeeakqkaaeatkvaeeakqkaa- 400
 QY 423 EFEPVPEQLETPISKRRRSRSKAKNLEKE 452
 DB 401 -----eatkvaeeakqkaaeatkvaee 423

RESULT 8
 AAY31954
 ID AAY31954 standard; Protein; 379 AA.
 XX
 AC AAY31954;
 XX
 DT 21-DEC-1999 (first entry)
 XX
 DE Glutathione-S-transferase:Weel(11-152) fusion protein.
 XX
 KW GST:Weel(11-152); glutathione-S-transferase; GST; Weel; human;
 KW mitosis; cell cycle checkpoint; cyclin-dependent kinase; Cds1.
 XX
 OS Chimeric - Schistosoma mansoni.
 OS Chimeric - Homo sapiens.
 XX
 PN W09949061-A1.
 XX
 PD 30-SEP-1999.
 XX
 PF 26-MAR-1999; 99WO-US06704.
 XX
 PR 27-MAR-1998; 98US-0079752.
 XX
 PA (SCRI) SCRIPPS RES INST.
 XX
 PI Russell P, Boddy MN, Funari B;
 XX
 DR WPT; 1999-591103/50.
 DR N-PSDB; AAZ20067.
 XX
 PT Truncated fusion proteins and related polynucleotides useful for
 PT screening for compounds which modulate Weel function
 XX
 PS Claim 6; Page 40-42; 46pp; English.
 XX
 CC The present sequence represents a fusion protein, termed
 CC GST:Weel(11-152), comprising glutathione-S-transferase and a
 CC truncated Weel protein composed of amino acid residues 11-152 of
 CC native Weel. The fusion protein was expressed in Escherichia
 CC coli cells that had been transformed with a pCEX vector into which
 CC a PCR fragment encoding amino acids 11-152 of Weel had been cloned.
 CC Weel participates in the regulation of cell cycle checkpoints during
 CC mitosis by catalysing phosphorylation of the cyclin-dependent
 CC kinase that initiates mitosis, Cdc2. Weel is itself phosphorylated
 CC by Cdc1. The fusion protein is useful in assays for evaluating Weel
 CC function and for screening for modulators of Weel and Cds1. It can
 CC be used in combination with other components of the cell cycle,
 CC particularly Cds1, Chk1, Cdc2, Cdc25 and other checkpoint proteins,
 CC in methods to identify and characterise regulatory pathways and to
 CC identify compounds which modulate components in the pathways,
 CC particularly Cdc1.
 XX

SQ Sequence 379 AA;

Query Match 28.1%; Score 1220; DB 20; Length 379;
 Best Local Similarity 60.3%; Pred. No. 6.2e-84;
 Matches 252; Conservative 32; Mismatches 50; Indels 84; Gaps 8;

QY 5 PILGYWKIKGLVQPRLLLEYLEEKYEELHYERDEGDKWRNKKFELGLEFPNLPYYIDGD 64
 DB 3 pilgywkikglvqprllleyleekyeelhyerdegdkwrnkkfelglefnpnyidgd 62

QY 65 VKLTQSMALIIRYADKHNMGLGCPKRAEISMLEGAVLDIRYGVSRAYSXDFETLKVDF 124
 DB 63 vltqsmaliiryadkhnmlgpcpkeraleismlegavldirygvsraysxdkfclkvdf 122

QY 125 LSKLPEMLKMFEDRLCHTKTYLNGDHYTHPDMLYDALDVLYMDPMLCLDAPPKLVCFKKR 184
 DB 123 lsklpemlkmfedrclchktylngdhvthpdmlydaldivlymdpmlcldapkvlvcfkk 182

QY 185 IEAIPQIDKYLKSSKYIAWPLQGWQATFGGDHPPKSDHLYVPRGSM----- 230
 DB 183 ieapqidkylksskyiawplqgwqatfggdhppksd-lvprgspgisgggghmrsv 241

QY 231 -LRLKRNQISKRIVFYLKQAKAFKGNHPCVSVCTITYSRPHCLDPTLKLSPMSSKT 289
 DB 242 glrrsqrsmlnra---cll-----aptpsslydannst 273

QY 290 TSLMPLQVNIAGNSFSAETPVLDLKKENETELANISGPHKKSTS----- 332
 DB 274 -----sstseqp-----ntsfslfgrkqtdtsspsfshaaplhlpsps 314

QY 333 -TSTRKARSKKATKATSVDPK----IDESVASYDSSTHLRSSRSKRPVNVSSSESE 386
 DB 315 fthsqpdiqavprpslfdprnlvrsrsrldgslspvaqvanpihtapspsd 372

RESULT 9

ID AAR81731 standard; Protein; 1140 AA.

AC AAR81731;

XX 29-MAY-1996 (first entry)

DT GST-SEP fusion protein fragment.

DE Glutathione-S-transferase; GST; FK506 binding protein; FKBP; PCR;

KW rapamycin; FKBP-rapamycin binding protein; Molt4 cell; amplify;

KW fusion protein; GST-FKBP12; immunomodulatory agent; primer;

KW antitumour agent; detection; antisense DNA; immune system.

XX Homo sapiens.

OS EP676471-A2.

PN 11-OCT-1995.

PD 07-MAR-1995; 95EP-0301475.

PF 13-FEB-1995; 9505-0384524.

PR 08-MAR-1994; 94US-0207975.

PR 26-SEP-1994; 94US-0312023.

XX (AMHP) AMERICAN HOME PROD CORP.

PA (UYCO) UNIV COLUMBIA NEW YORK.

PI Caggiano TJ, Chen Y, Failli AA, Molnar-Kimber KL;

PI Nakanishi K;

PI WPI; 1995-346091/45.

DR N-PSDB; AAT00771.

XX New effector proteins of rapamycin - which bind to a

PT glutathione-S-transferase-FK506 binding protein-rapamycin complex

XX Example 2; Page 29-35; 44pp; English.

PS This sequence represents a fragment of a fusion protein between

CC glutathione S transferase (GST) and sirilimus effector protein (SEP)

CC gene. It represents the fragment beginning at the linker sequence

CC between the GST and SEP45 fragment. The cDNA encoding this sequence

CC was isolated from Molt 4 human T-cell leukaemia cells (ATCC CRL 1582).

CC The sep gene was amplified in five fragments using the primers given in

CC AAT00756-69. The amplified SEP gene was inserted into a vector already

CC containing the GST gene and the fusion protein expressed. The fusion

CC protein was used in the isolation of a protein of mammalian origin

CC which binds a glutathione-S-transferase (GST)-FK506 binding protein

CC (FKBP)-rapamycin complex. The FKBP-rapamycin binding protein and

CC corresponding DNA was isolated from Molt4 cells using a complex of the

CC fusion protein GST-FKBP12 and rapamycin. The isolated proteins have

CC molecular weights of 125, 148, 208 and 210 kD. They can be used for

CC identifying an immunomodulatory, or an antitumour agent. They can also

CC be used in the detection of rapamycin, rapamycin analogues or

CC metabolites when complexed with FKBP. Antisense DNA can be used to

CC modulate the immune system of a mammal.

XX Sequence 1140 AA;

SQ

Query Match 28.0%; Score 1217.5; DB 16; Length 1140;
 Best Local Similarity 36.2%; Pred. No. 4.8e-83;
 Matches 344; Conservative 103; Mismatches 261; Indels 241; Gaps 34;

QY 5 PILGYWKIKGLVQPRLLLEYLEEKYEELHYERDEGDKWRNKKFELGLEFPNLPYYIDGD 64
 DB 3 pilgywkikglvqprllleyleekyeelhyerdegdkwrnkkfelglefnpnyidgd 62

QY 65 VKLTQSMALIIRYADKHNMGLGCPKRAEISMLEGAVLDIRYGVSRAYSXDFETLKVDF 124
 DB 63 vltqsmaliiryadkhnmlgpcpkeraleismlegavldirygvsraysxdkfclkvdf 122

QY 125 LSKLPEMLKMFEDRLCHTKTYLNGDHYTHPDMLYDALDVLYMDPMLCLDAPPKLVCFKKR 184
 DB 123 lsklpemlkmfedrclchktylngdhvthpdmlydaldivlymdpmlcldapkvlvcfkk 182

QY 185 IEAIPQIDKYLKSSKYIAWPLQGWQATFGGDHPPKSDHLYVPRGSMRLKLKNIQISKRI 244
 DB 183 ieapqidkylksskyiawplqgwqatfggdhppksd-lvprg-----spdisggg 234

QY 245 VFTILQK-AFKGNHPCVSVCTITYSRPHCLDPTLKLSPMSSKTTL-----SMLPQVN 298
 DB 235 gggildsdmsfkyaslcgks-----grlalahktlivilgvdpsrqldhplptvhpqvt 288

QY 299 IG--ANSFSAETPVD-----LKKENETELANISGPHK----- 328
 DB 289 yaymkumwksarkidafqhmghfvqcmggaqhaiaiedqgkqlhklmarcfklkgew 348

QY 329 ----KSTSTSTRKRARSSKKKATDSVSKIDES-----VASYDSSTHLRSSR--- 372
 DB 349 qinlqginestipkvlygysaateh-----drswykwahawamnmfeavllkhhqndprd 403

QY 373 SKKPNVNVSSSESEEEQISKATKKVKQKEEBEYVEEVDKSLKNESSDEFEVYVPEQL 432
 DB 404 ekklrlrhasga-----nitnattaattaattastegsnseaeastensp-tpspl 456

QY 433 ETPISK-----RRRSKSAKNLE---KESTMILD-DHAP----- 462
 DB 457 qkkviedlskllmytvpavqgffrsislsrgnnlqdtlrvltlwdyghpvdvneale 516

QY 463 -----REMFCDLDPKIPWRGRGLGYACLNITLRSMKERVCSRTC---RIIT 505
 DB 517 gvkaiqidtwlqvipqliaridprlvgrlihqldtdgrhyhpaqialyptlvasksttt 576

QY 506 IORDGLESVKQLGTQNVLDLKLVEVNNHNFGLHFMWRVSSDLFPFAS--HAKYGYTLEFAQ 563
 DB 577 arhnaankl-----lkmcelhsltlvqamvseelirvailwh-----emwh 619

RESULT 13
 AAB70883
 ID AAB70883 standard; Protein; 856 AA.
 AC AAB70883;
 DT 12-JUL-2001 (first entry)
 XX Expression vector pGEX-4T-1-AmSu5 containing amylosucrase.
 DE Amylosucrase; EC 2.4.1.4; fusion protein; GST; glutathione-S-transferase;
 KW poly(1,4-alpha-glucan); film production; food additive; cyclodextrin.
 XX Unidentified.
 OS WO200125449-A2.
 PN 12-APR-2001.
 PD 04-OCT-2000; 2000WO-EP09695.
 PF 07-OCT-1999; 99DE-1048408.
 PR (AXIV-) AXIVA GMBH.
 PA Bengs H, Polakowski T, Held A, Gallert K;
 PI WPI; 2001-328330/34.
 DR N-PSDB; AAF61712.
 XX Amylosucrase immobilized as fusion protein with anchoring sequence,
 PT useful in continuous preparation of poly(1,4-alpha-glucan) from sucrose
 PT
 PT
 XX
 PS Claim 2; Page 34-38; 38pp; German.
 XX
 XX This invention describes a novel amylosucrase (AS), immobilizable on a
 CC solid phase, which comprises a fusion protein (FP) of functional units
 CC of AS, an anchoring sequence, and optionally additional auxiliary
 CC sequences. The invention also describes (1) nucleic acid (I) encoding
 CC FP; (2) expression vector containing (I) and able to express FP in a
 CC host cell; Escherichia coli containing the vector of (2); (3) anchoring
 CC sequence, or its functional variants or fragments, of at least 8
 CC nucleotides that encodes an epitope, a high-affinity binding partner or
 CC GST (glutathione-S-transferase); (4) solid phase for immobilizing AS
 CC comprising glutathione-Sepharose; (5) combined, stable catalyst (A)
 CC comprising FP immobilized on Sepharose for production of
 CC poly(1,4-alpha-glucan) (II); and (6) biocatalytic production of (II).
 CC AS is used in production of poly(1,4-alpha-glucans) (II), useful for
 CC producing films, as food additives, as starting materials for
 CC cyclodextrins and as auxiliaries in pharmaceutical formulations.
 CC Immobilized AS makes possible efficient, inexpensive and continuous
 CC production of poly(1,4-alpha-glucans) (II), and it can be used
 CC repeatedly. Compared with known methods, specificity is improved
 CC (increased yield of (II) and reduced formation of patainose) and
 CC reaction is complete within 24 hours, compared to 48-72 hours for batch
 CC methods. This sequence represents the expression vector construct
 CC pGEX-4T-1-AmSu5 which contains the amylosucrase (EC 2.4.1.4) described in
 CC the invention.
 XX
 XX Sequence 856 AA;
 XX
 Query Match 27.8%; Score 1208.5; DB 22; Length 856;
 Best Local Similarity 39.0%; Pred. NO. 1.5e-82;
 Matches 331; Conservative 75; Mismatches 239; Indels 203; Gaps 29;
 QY 5 PILGYWKIKGLVPTRLLEYLEEKYEHHLYERDEGDKRNKKFELGFLFPNLPYYIDGD 64
 DB 3 PILGYWKIKGLVPTRLLEYLEEKYEHHLYERDEGDKRNKKFELGFLFPNLPYYIDGD 62
 QY 65 VKLTQSMALIRYIADKHMLGGCPKRAEISMLEGAVLDTIRYGVSRVAYSKDFTLKVD 124

Db 63 vkltqsmaliiryiadkhmlggcpkeraeismlegavldirvgvsriayskdfetlkvdf 122
 QY 125 LSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVLVYMDPCLDAFPKLVCFKKR 184
 Db 123 lsklpemlkmfeurichktylmgdhvthpdfmlydalvlymdpmcldafpklvcfkk 182
 QY 185 IEATPQIDKYLKSSKYIANPLQGWQATFGGDDHPPKSDHLVPRGS-----MLRLKXNI-- 238
 Db 183 iealpqidkylksskylawplqgwqatfggddhppksd-lvprgsbefmlqyikrlid 241
 QY 239 -----QTSKRIVFTILKQKAFKGNHPCVPVSVCTITYSRFHCLPDTLKS 281
 Db 242 iytpeqragiekedwrgfsfrm-----dthfplmmeldsvyngneallpmlm 291
 QY 282 LLLPMSSKTTLSMLPQVNIIGANSPSAETPDLKKENETE--LAN--ISGPHKSTSTSTTRK 337
 Db 292 llaqawqs---ysqrn-----sslkldiatenpndwilsnkqvggcyvdlfagdlk 341
 QY 338 RAKSSKKKATDSVSDKI---DESVASYDSSSTHLRRSRKPKVNYNSSSESE----- 386
 Db 342 -----gldkklpyfqlgltylhlmpfkpegksdgyavssyrdvnpalgti 390
 QY 387 -----SEQISKATKKVKQKEEYVEEDVDEKSLKNSSSDEFEFVPEQLETP 436
 Db 391 gdlreviaalheagisavvdfifnhtsneh--ewagrcagagplfdnfyfipdr-rmpd 447
 QY 437 SKRRSRSSAKNLEKSTWMLDDHAPREMFDCLDKPIPWGRGLGYACLN----- 486
 Db 448 qydrtlreifpdqhgpgsgled-----grwvttfnsgfdwnlmysnp 490
 QY 487 -ILRSMKRVFCSTCRITTIQDGLFSV-KQLCT--QNVLDLILKIVE-WNNHFGIHFMR 541
 Db 491 wvframagmflanlgvdlrmdavafwkqmgtsenlpqahalirafn-----avmr 545
 QY 542 VSSDLFPFASHA-----KGY-----TLEFAQSHLEEVCKL--ANKYNH 578
 Db 546 iaapavfikeaivhpqvvgygqdcqigynplqmallwntlatrevnlhqlalyrh 605
 QY 579 RLTHWPGQYQIASPRVWVDSALRLDAY-----HDEILSRMKLNEQLNKDAVLIHL 631
 Db 606 nlpehtawnyvrshddigwtfdadaaylgisgydhrqfnnrfvnr----- 653
 QY 632 GGTFEGKKEITLDRFNKYQ-----RLSDSVKARLVLENDVSWVQDLLPLCQELNPLVL 687
 Db 654 ---fdgsfargvpfyqngstgdcvsgtaaalvglaqddp--havdrkilysl----- 702
 QY 688 DWHHINIVPGTLREGSLDLMLP-----IPTIRE-----TWTRKGITOKQHYSES 731
 Db 703 -----alstggipilygdevgtlndddwsqdsksddsrwahrpnynealyaq 752
 QY 732 ADPTAISG 739
 Db 753 ndpstaag 760
 RESULT 14
 AAR79948
 ID AAR79948 standard; Protein; 362 AA.
 AC AAR79948;
 XX
 XX 26-FEB-1996 (first entry)
 DT
 XX
 DE G17 antigen.
 XX
 KW Syphilis; G15 antigen; membrane antigen; glutathione-S-transferase;
 KW GST; fusion protein; membrane antigen; diagnosis.
 XX
 OS Chimeric Treponema pallidum;
 OS Chimeric unidentified bacterium.
 XX
 FH Key Location/Qualifiers
 FT Protein 1..228

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	3112	100.0	599	3	Q10988	schizosacch	
2	1011.5	32.5	656	3	Q01408	neurospora	
3	359	11.5	317	2	Q9L4F0	bacillus ce	
4	359	11.5	325	2	Q9L4E9	bacillus ce	
5	306.5	9.8	322	2	Q9K9P8	bacillus ha	
6	229.5	7.4	305	2	Q9S0W9	deinococcus	
7	229.5	7.4	326	2	Q9RTE6	deinococcus	
8	164	5.3	1588	11	Q9ESK9	mus musculu	
9	156.5	5.0	368	5	Q45198	caenorhabdi	
10	155	5.0	1641	5	Q9GRZ3	caenorhabdi	
11	154.5	5.0	782	5	Q26104	plasmodium	
12	149.5	4.8	782	5	Q9U414	plasmodium	
13	148.5	4.8	911	3	Q06704	saccharomyc	
14	148.5	4.8	1359	5	Q02061	caenorhabdi	
15	148	4.8	1359	5	Q9U7E0	caenorhabdi	
16	147.5	4.7	782	5	Q26007	plasmodium	
17	145.5	4.7	401	3	Q13885	schizosacch	
18	145.5	4.7	782	5	Q9U430	plasmodium	
19	145	4.7	1957	5	Q04010	onchocerca	

Db 276 KPFLIIIAKKNHNFDMIESKQKDIALFQL 305

RESULT 5

Q9K9P8 PRELIMINARY; PRT; 322 AA.

AC Q9K9P8;

DT 01-OCT-2000 (Tremblrel. 15, Created)

DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)

DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)

DE BH2597 PROTEIN.

GN BH2597.

OS Bacillus halodurans.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI_TaxID=86665;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C-125 / JCM 9153;

RX MEDLINE=20512582; PubMed=1058132;

RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,

RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,

RA Horikoshi K.;

RT *Complete genome sequence of the alkaliphilic bacterium Bacillus

RT halodurans and genomic sequence comparison with Bacillus subtilis.;

RL Nucleic Acids Res. 28:4317-4331(2000).

DR EMBL: AP001516; BAF06316.1; -.

KW Complete proteome.

SQ SEQUENCE 322 AA; 36900 MW; 4B0DC9291CD1E7B4 CRC64;

Query Match 9.8%; Score 306.5; DB 2; Length 322;

Best Local Similarity 26.6%; Pred. No. 5.3e-12;

Matches 85; Conservative 75; Mismatches 123; Indels 37; Gaps

Qy 247 RGRIGYACLTILSRMKRVFCSTCRITTIQR-----DGLSVKQLGTGNVLDLKLVE 301

Db 2 RIQGYVAMSMELANASP---SKTWTATQFEKIEDHEAGRLKERTAKTNLHNCRLRLK 57

Qy 302 WNNHFGTHFRVSSDLPPFASH---AKYGYTLEFAQSHLEEVCKLANKYNHRLTMHPGOY 358

Db 58 HNLAYQISFFRLSKSLPLVNNHPLTEGKYLELAIAE-ELQAVGEFASEHOMRIDFPHDF 116

Qy 359 QTASPREVVVDSAIRDLAYHDEILSRMKLNEQLNKDAVLIIHLGTFEGCKKETLDRFRK 418

Db 117 VVLNSEAKEITRRSLQTLHYKLLKGWIDPRHR---CVLHVGGKKKGVEAGLEQFIE 172

Qy 419 NYQRLSDSVKARLVLENDVSVSWSDLLPICOELNPLVLDMHHHNIIVPTGLREGSLDLM 478

Db 173 NTASIPKSLLSMILENDKDSYTIIDVLYLGEKLAIPVLDIHHDVL-----HRSKSLQ 227

Qy 479 PLIFTIRETWTKGITOKHYSESADPTATSGMKRAH-----SDRVFDF-----PCDP 528

Db 228 ETWQRIATVEDSPVPKIIHLS-----SPLSGEDPRHHDYINADRIAFLEHIGDAVD 282

Qy 529 TMDLMIEAKEQEAVFELCR 548

Db 283 HLHVWIEAKKKDIALFOLMK 302

RESULT 6

Q9S0M9 PRELIMINARY; PRT; 305 AA.

AC Q9S0M9;

DT 01-MAY-2000 (Tremblrel. 13, Created)

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)

DE UV-ENDONUCLEASE.

GN UVSCDE.

OS Deinococcus radiodurans.

OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.

OX NCBI_TaxID=1299;

RN [1]


```
QY 198 VVPEQLETPI-----SKRRSSAKN-----LEKESTMNLDHAPREM 236
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 754 QSPMMVESLYSSVINAIDSRQDSTRNGEGDRAALHVLQEKRAAQQDSHTSIQT 813
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 237 FD---CLDKPIPNRGLGYA-----CLNTILSRKWERFVCSRTCRITTIQDGLSVKQLG 289
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 814 IKDQLCHFTFVQKQCDLANYLKCTAVEIRNIIKVKCSLEITLTKKHQOQLSLK-IE 872
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 290 TONVLD-LIKLVENHNHFGIHFWRVSSDLPPFASHAKYGYTLFEAOSHLEEVCKLANKYN 348
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 873 YECKLDALVQDSENVN---KILKLENLV-----SUEEA--LQNKDN 910
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 349 HRLTMHPGOYTOIASPREVVV-----DSAIRDL- 376
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 911 -----EFTSIKHKDAIVCQOEKQDKLLEWETIMHTQHCIEIKELQSRMALEDLK 962
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 377 AYHDEILSRMK-----LNEQLNKDVLIIHLGOTFGCK-----ETLDRFRK-NYORLS 424
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 963 KLHDEKTESRAEFQCLEENHLKLEEDTLHIRTQEFKVTMDHNSLEKLEKKNQOQRI 1022
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 425 DSVK--ARLVLEND---DYSWSVQDLPLCQELNIPLVLDWHHNIIVPCTLRGSLDLM 478
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1023 QMLSESHASTQOEKQQLKLVLSLSDMRCKLEVEL-----ALKEAETDEM 1070
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 479 PLIPTIRETWRTKGITOKHYSBESAD--PTAISGMKRRRAHSDRVDFPPCDPTMDLMIEA 536
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1071 KIL--LEESRTQOEKMLKSLLEQETENLRTEISKLNQKIHNNESYQVGLSELRALM--T 1126
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 537 KEKEQAVFELCRRYEIQNPCCPLIEIMGPEVDQ-----RDGYPPGAEKRLTAR-----KRR 588
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1127 IEKQDCISILSRHEES-----NILKAELDNVTSLHQAY---ETIEKKLKEQIVELQTR 1178
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 589 SRKE-----EVEEDEK 599
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1179 LNSELSALEKQKDEK 1193
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 9
ID 045198 PRELIMINARY; PRT; 368 AA.
AC 045198;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DE 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
GN W09G12.7 PROTEIN.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Lalster N., Latreille P.,
RA Lighning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.:
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Beck C., Fronick W., Wilson R.;
RA Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
```

```
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF047663; AAC04447.1; -.
SQ SEQUENCE 368 AA; 41823 MW; AE74C475F8340CDA CRC64;

Query Match 5.0%; Score 156.5; DB 5; Length 368;
Best Local Similarity 26.1%; Pred. No. 0.021;
Matches 57; Conservative 37; Mismatches 61; Indels 63; Gaps 8;

QY 73 NGFSAETPVDLKK-----ENETELANISGPHKKSTSTSTRKRARGSKKKKATDSVSK 124
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 85 NDSSESSDDDKKKKKKKKKSSDSSDSDGKKKKKKKKSSDSSDSDS 144
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 125 IDSVASVDSSTHRRSRSSKPKVNYNSSS-----E 157
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 145 SDSSDSSDSDDKKKKKKKKKSGSSDSSDSDSDDEKPKKKKKSKPKKAPK 204
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 158 SEQIIS-KATKVKQKEE-----EYVEEVEDEKSLKNSSSDEFE----- 196
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 205 SDEEYTIKTIKKPKKPEDLERRRHVRVKKTEHEEEKD-KEEHESNDENMEFDTARGA 263
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 197 PVVPEQLETPISKR-RRSRSAKMLEKESTMNLDHAP 233
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 264 PSSPK-----LTRRAKRSESSDDEVELOKTPIDDNKP 296
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 10
QY 09GRZ3 PRELIMINARY; PRT; 1641 AA.
AC 09GRZ3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
GN Y59A8B.1 PROTEIN.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Sulston J.E.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; AL132898; CAC14406.1; -.
DR InterPro; IPR002965; PrichExtensn.
DR PRINTS; PR01217; PRICHEXTENS.
SQ SEQUENCE 1641 AA; 181927 MW; F6C38533AA158D1D CRC64;

Query Match 5.0%; Score 155; DB 5; Length 1641;
Best Local Similarity 19.2%; Pred. No. 0.17;
Matches 126; Conservative 91; Mismatches 227; Indels 212; Gaps 29;

QY 55 PMSSTKTTLSML--POVNTGANSFSAETPVDLKKENETELANISGPHKKSTSTST-----R 107
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 841 PVVTKAALKMVGPRKKTGRRR-----KKDRGASPSPPSPPPPTTLTPIVLPR 891
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 108 KRARSSKKKAT-----DSVSDKIDESVASYDSSSTHLRRSRSSKPKVNY 150
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 892 KKORIEKKKLTTPPPQQAQPVTSHTPPVPEQLMSRRKKQIMMERHSLDHIQ-----FKLIEI 946
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 151 NSSSESESEQISKATKKVKQKEEYVEEVEDEKS-----LKNSSSDEF 195
  || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```

Db 947 ELAAKRRKAAEAAAARAAVEQKDEK--AEVENRETTPGPSTSMRSSLHTPNTSEDEV 1004
QY 196 EPVPEQLETP-----ISRRSRSSAKNLEKSTW-----226
Db 1005 IFVEPSTLEKPERNGTTEERVMTADQAFMDAKIEEARRSRMSTRDCSVVSTLGPVKSK 1064
QY 227 ---NLDD--HAPREMFCDLKPWPGRGL-----GYACLN 256
Db 1065 ASQRLHDIIEGKEELEDSDMPTNNGTLAGILYPMRSEAEVSSNHRSEGAGGMSLK 1124
QY 257 TILSRMKERVCSCTRTITQDGLSEVKGQTONVL---DLIKLVE-----WNNHFG 307
Db 1125 HHLARKNELKEANVARSEILKAVVRORREIGVPTTILMSKSAIELVEDEKERKHNKN 1184
QY 308 -----IHPMRVSSDLFPFASHAKYGYLEFAQSHLE-----EVGKLANKYNH---RLTMH 354
Db 1185 KTLSPHDYVRSKNEAEKAEFGK--GGTMRITNRNKLMLTRQFDLPKMSRFRKRVRIIRH 1243
QY 355 PG-----QYTOI---ASPREV-----VVDSATRDLAYH 379
Db 1244 PNGMATIISCYNOIKQHLGNEMKHFQRFVRLGFAENNGVPLFAIGVMENAAE--ALH 1301
QY 380 DEI-----LSRMKLNQELNKDAVLIHILGGTFEGKKTLDLRFKKNYQRLSDSVKARL 431
Db 1302 DQFELAKNSPNTQVKGSLTNKQFIETPMKYYESAMETLDMGTFRFGPLMS--LSMV 1359
QY 432 VLENDVSVSQDILLPCQELN-----IPLVLDWHHHNIVPCTLRGSLDMLPLIPIRET 487
Db 1360 GTKNEEAGNGKEML---DALNAAPFLGTPMPWGDSEVOG--IKEDSDGQIP-----1409
QY 488 WTRKG-----ITOKHYESADPTAISGMKRR-----AHSRDFVDFPPCDPT 529
Db 1410 WVRPGQWVPTDGNRSTPRHPLATRGNDRRETAFTDRTNAHADQVRESTEDDPT 1465

RESULT 11
ID Q26104 PRELIMINARY; PRT: 782 AA.
AC Q26104;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE RHOPTRY ASSOCIATED PROTEIN 1.
GN RAP-1.
OS Plasmodium reichenowi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5854;
RN [1]
RP SEQUENCE FROM N.A.
RA Shi Y., Lal A.A.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20986; AAA63682.1; -.
SQ SEQUENCE 782 AA: 90160 MW; DE6D1BE2FAC308A9 CRC64;

```

Query Match 5.0%; Score 154.5; DB 5; Length 782;
 Best Local Similarity 20.9%; Pred. No. 0.074;
 Matches 147; Conservative 94; Mismatches 240; Indels 221; Gaps 37;

```

QY 43 FHCLPDTLKLSPSSKTTLSMLPQVNIIGANSFSAETPVDLKKENETELANISGPHKST 102
Db 122 FKASPSVVKTTPTPGTHTS-----GSKSSS---PSSTKSSPSNVKTAS--PHGESN 168
QY 103 STSTRKRARRSKKATDSVDKIDESVAS-----YDSSTHLRSSRS-----144
Db 169 S--SEESTSKSRKSASVGIAGADEEVPAPKNLTLPLEELYPTNVNLFNYKSLNME 227
QY 145 -----KKPNYNSSSESESEIOISKATKKVKQKEEEVEEVEDEKSLKNE- 189
Db 228 NINILKNGQLVAQKEEFDYDENMEKAKEK--QKALEKIGETDEEFPMTDDKFLNQV 286
QY 190 -----SSSDEFEFVVP-----EOLE--TPISKRARRSRSSAKNLEKSTW-----NLDHHA 232

```

```

Db 287 KERNVAGSRFRFSKLNLPFKKDEVTEKTEVSKRTYSGTGNVNDREAKILGVGATYQYEP 346
QY 233 PREMFCDLKPWPGRGLGYACNLILSRMKERV--FGSFTCRITTIQDGLSEYKQLGT 290
Db 347 ETMLYNCPPN-----SHLFDITSSLOGKLDIKKRENMISTTFEQOKECKLNMG- 395
QY 291 QNVLDL-IKLVENH-----NFGIHFMRV---SSDLFPFASHAKY-----GYTLE---F 332
Db 396 --VLDLELKDTECKFGTCIGSRGEHHLRLYEFDLFPHPNIDYLTTLADGYKQKNNHIY 453
QY 333 AQSH-----LEEVGKLAN-----KYNHR-----LTMHPGQ--YTIQI 361
Db 454 ELSHVNFCLLNPKTLEEFLLKKEIKDLMGDDLLIKYKFNDFNFMSSIT'CHIESLIYDDI 513
QY 362 ASPREV-----VDSAI---RDLAYH-----DELSRMKL---NEOLN 393
Db 514 EASQDVAALVKTAKSKLHVITSLGSKARKLVYKIYSEIQKNPDDELYEKLTIWIYDIYML 573
QY 394 KDAVLIIHGGT---PEGKKTLDLRFKKNYQRLSDSVK-----ARLVLEN---DDVSWS 441
Db 574 KRYTAYALEGVCSTYENDKTOMDTLHLYNKLVDSRVYSSCFKNVIYVNAIMSGIHEK 633
QY 442 VODLPLCQELNIPVLVDWHHHNIVPCTLRGSLDMLPLIPIRETWT'KGTQKHYS 501
Db 634 IKHFLKLVPRQN--FLLDYHNSIF-----EKEIKPVVKYST 668
QY 502 S---ADPTAIS---GMKBRASDRVDFPPCDPTMDLMLAEAKEQVAFELCRRYELON 554
Db 669 SHVYDPTVASVAYNLDRTWT'ITINDY-----FEAKRKELTIVSRMKTDM-- 716
QY 555 PCPPEIMGPEYDQTRDGYYPGPAEKRLTARKRRSRKKEVEE 596
Db 717 ----LSQNDSEKIPND----KSANSKLATRLMRKFAEIRD 750

RESULT 12
ID Q9U414 PRELIMINARY; PRT: 782 AA.
AC Q9U414;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE RHOPTRY ASSOCIATED PROTEIN-1.
GN RAP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Li X.R., Luo S.H., Yu X.B., Shan Z.X., Ma C.L.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF206631; AAF15365.1; -.
SQ SEQUENCE 782 AA: 90082 MW; 8EIF4CF2883903FD CRC64;

```

Query Match 4.8%; Score 149.5; DB 5; Length 782;
 Best Local Similarity 21.0%; Pred. No. 0.15;
 Matches 147; Conservative 94; Mismatches 241; Indels 219; Gaps 39;

```

QY 43 FHCLPDTLKLSPMSKTTLSMLPQVNIIGANSFSAETPVDLKKENETELANISGPHKST 102
Db 122 FKASPSVVKTTPTPGTQT-----SGLKSSPSSTKSSPSNVKAS--PHGESN 168
QY 103 STSTRKRARRSKKATDSVDKIDESV---ASYDSSTHLRSSRSRKKPVNYN---SSSES 156
Db 169 S--SEESTTKSKSRKSASVAGIAGADEAPAPKNTLTP---EELYPTNVNLFNYKYSLN 223
QY 157 ESEEQISKATKK---VKQKEEEVEEVD-----EKSLLKNSSSDSEPEPVPEQ---LE 204
Db 224 NMEENITLKNEGDLVAQKEEPEYDENMEKAKODKKALEKTGKOSDEEPPMFSENKFELE 283

```



```

OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Favello A., Fulton L.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierri-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX STRAIN=BRISTOL N2;
RA Fulton R., Wohlmann P.;
RT "The sequence of C. elegans cosmid B0041.";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO HELICASE C-TERMINAL DOMAIN.
DR EMBL; AF000196; AAC24256.1; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR000330; SNF2_N.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00176; SNF2_N; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
KW ATP-binding; Helicase.
SQ SEQUENCE 1359 AA; 156235 MW; 5F2ACE099EDED4F4 CRC64;

Query Match 4.8%; Score 148.5; DB 5; Length 1359;
Best Local Similarity 20.9%; Pred. No. 0.36;
Matches 136; Conservative 96; Mismatches 222; Indels 197; Gaps 34;

QY 83 LKKNETELANTISGPHKKSTSTSTRKRARSSKKKATDSVSKIDSVASVDSSTHLRSS 142
DB 229 VKKESE---DEAEKK--KTEKKRKSTSESESE--SEKSDDEEEKSSP-----KP 277

QY 143 RSKKPVNYS--SESESEEQ-----ISKATKKVKQKEEEVEEVDKKS 185
DB 278 KKKKPLAVKKLSDESESDVEVLPOKKRGAVTLISDSEKDKQKGESE--ASDVEKV 336

QY 186 LKNESSEDEFEPVPEOLETPTISKRRSSRSKAKNLEKSTWNLD-----DHAPR 234
DB 337 SKKAKKQSSGSDSGSTIVNRKSKKKKPEKKRKGIIIMDSKSKLOKETIDAEK 396

QY 235 EMFDCIDKPIWGRGLGYACLNTI-----LRSMKRVFCRTRITTIQDGL----- 283
DB 397 ERRKRLEKK-----QKEFNGVLEEGEDLTEMLTGTSSQKLSVVLDPDSVDEE 448

QY 284 SVKQLGTQNVLDLIKLVENHNFHFMVRSDDLFPFASHAKYGYTFEFAQ----- 334
DB 449 SKKPEVHN--SLVRILKPHQAHQIOFM-----YDCAFESLDRLDTGEGS 490

QY 335 -----SHLEEVGKLAN--KYNHRLTWRP--GOYQ---IASPREVVVD---SAIRLAYHD 380

```

```

DB 491 GGILAHCMGLGKTLQVITELHTVLMHEKIGEKCKRVLVVVVKNVILNFKEFQKWLVDND 550
QY 381 EILSRMKLNELQNKDAVLIHILGGTFEGKKETL-----DRFR----- 417
DB 551 EELDTIDVNE-----LDSYKTIEDRRKALKAWHSSKTPSVMIIGYDLFRILTVEDD 601
QY 418 -----KNYQRLSDSVKARL-----VLENDVSVSWQDILLPLCOELN 453
DB 602 PKKKPKNKRNRLEKAKEDFRKYLQNGPDMVVCDEAHKLNKDDLSALS--KCMVKILTKRR 660
QY 454 I-----PLVDWHHHNIVPGTLREGSLDMLPLIPTIRETWTRKGITQKOHYSESADPTAI 508
DB 661 ICLTGTPLQNNLMYHCHMVNFKPG-----LLGKTFEAFNRFVNIINRGRTKADSPLEV 714
QY 509 SGMKRRAHSDRVFD-FPPCDPTMD--LMTAEKKEQAVFELCRRYELQNPCCPL----- 559
DB 715 SFMKRRCHV--LYDHLKCKVDRKDRVLTETAPPKOEYVINVRQTERQ---CALYNALFN 769
QY 560 ETMG-----PEYDO--TRDGYPPGA---EKLRTARKRSRKEEVED 597
DB 770 DIVGDSGLSKRLLPDYHMFSTRITWHPYQLVLHEQRM--ERERVMDREDAEE 818

RESULT 15
Q9U7E0 PRELIMINARY; PRT: 1359 AA.
AC Q9U7E0;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE XNP-1.
GN XNP-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Telodoridae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99365296; PubMed=10433961;
RA Villard L., Fontes M., Ewbank J.J.;
RT "Characterization of xnp-1, a Caenorhabditis elegans gene similar to
RT the human XNP/ATR-X gene.";
RL Gene 236:13-19(1999).
CC -!- SIMILARITY: TO HELICASE C-TERMINAL DOMAIN.
DR EMBL; AF134186; AAD55361.1; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR000330; SNF2_N.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00176; SNF2_N; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
KW ATP-binding; Helicase.
SQ SEQUENCE 1359 AA; 156191 MW; EB4342547D4F4E64 CRC64;

Query Match 4.8%; Score 148; DB 5; Length 1359;
Best Local Similarity 27.1%; Pred. No. 0.38;
Matches 52; Conservative 29; Mismatches 63; Indels 48; Gaps 7;

QY 84 KKNETELANTISGPHKKSTSTSTRKRARSSKKKATDSVSKIDSVASVDSSTHLRR 140
DB 87 KSESESEDEEDRKKSKKKKVDQKKKKKKRTTSSSEDEDEDEEQSK---KK 143

QY 141 SSRKPKPVNYSSESESEEQI--SKATKKVKQKEEEVEEVEEVE-----KS 185
DB 144 SKKTTKOTSSSESESEDEERKVKKKKKNKKEKSVKRAETSESEDEDEKPSKSKGLKK 203

QY 186 LKNESSED-----EFEPVPEOLETPTISKRRSSRSKAKNLEK 223
DB 204 AKSESESEDEKVKKKKKKVKKKKSESEDEDEAPEKKKT--EKKSKTS---EES 258
QY 224 STMNLDDHAPRE 235

```

Db 259 SESEKSDDEEEE 270

Search completed: January 15, 2002, 14:06:13
Job time: 952 sec



Db 241 DKPIPWGRGLGYACLNTILRSKMERVFCSTCTRTTIQROGLSVKQOLGTONVLDLKLK 300
QY 301 EWNHNFGEHMRVSSDLPPFASHAKYGYTLLEFAOHLBEVGLKANKYNHRLTMHPGQYTO 360
Db 301 EWNHNFGEHMRVSSDLPPFASHAKYGYTLLEFAOHLBEVGLKANKYNHRLTMHPGQYTO 360
QY 361 IASPREVVDSAIRDLAYHDEILSRMKLNEQLNKDAVLIIHLGGTFEGCKKETLDRFRKNY 420
Db 361 IASPREVVDSAIRDLAYHDEILSRMKLNEQLNKDAVLIIHLGGTFEGCKKETLDRFRKNY 420
QY 421 QRLSDSVKARLVLENDVSWSVQDLPLCQELNIPVLVDWHHNTVPGTLREGSLDLMLPL 480
Db 421 QRLSDSVKARLVLENDVSWSVQDLPLCQELNIPVLVDWHHNTVPGTLREGSLDLMLPL 480
QY 481 IPTIRETWKGIITOKOHSYSESADPTAISGMKRRRAHSRVDFFPCDPDTMDLMEAKEKE 540
Db 481 IPTIRETWKGIITOKOHSYSESADPTAISGMKRRRAHSRVDFFPCDPDTMDLMEAKEKE 540
QY 541 QAVFELCRRYELQNPCCPLEIMGPYDQTRDGYPPGAEKRLTARKRRSRKEEVEEDEK 599
Db 541 QAVFELCRRYELQNPCCPLEIMGPYDQTRDGYPPGAEKRLTARKRRSRKEEVEEDEK 599
RESULT 2
S55262
UV-endonuclease - Neurospora crassa
C:Species: Neurospora crassa
C:Date: 10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 20-Jun-2000
C:Accession: S55262
R:Yajima, H.; Takao, M.; Yasuhira, S.; Zhao, J.H.; Ishii, C.; Inoue, H.; Yasui, A.
EMBO J. 14, 2393-2399, 1995
A:Title: A eukaryotic gene encoding an endonuclease that specifically repairs DNA damage
A:Reference number: S55262; MUID:95292980
A:Accession: S55262
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-656 <YAJ>
A:Cross-references: EMBL:D11392; NID:g1526560; PID:g927215
C:Genetics:
A:Gene: uvel

Query Match 32.5%; Score 1011.5; DB 2; Length 656;
Best Local Similarity 36.6%; Pred. No. 3.2e-48;
Matches 241; Conservative 88; Mismatches 149; Indels 181; Gaps 18;

QY 110 ARSKKKATDSVSKTIDESVASYDSST-----HLRSSR-----SKKPVNYSSESSE- 157
Db 3 SRKSKAAALDTPQSESTSTSSLDSSAPSPARNLRSGRNILQPSSEKDRDHRKSGEL 62
QY 158 -----SEBQISKATKKVKQKEEYVEEVEDEK 184
Db 63 AGRMGKDHCHLREGEKEQGVKMAIEGLARMERLQKQKQLEED----- 114
QY 185 SLKNSSDEFPVPEQLETPISKRRS-----RSSANKLEKSTMNLDHIA----- 232
Db 115 GIPVPSVSRF-PTAPYHKKSTNAEEREAKEPVLKTHSKDVEREAIEGVDVVVKMEPAAT 173
QY 233 -----PREMFCDLCK-----PIPWGRGLGYACLNTILRSKMERVFCSTCTRTIT 275
Db 174 NIEPEDAQDAERGAARPPAVNNSYLPWPKGRGLGYACLNTYLRNSKPPFSSRTRCMA 233
QY 276 TI-----ORGLSVKQOLGTONVLDLKLIVENHNFG 307
Db 234 SIVDHRHPLQFEDEPEHLKKNPKDSKEPODELGHKFVQELGLANARDIVKMLCNKRYG 293
QY 308 IHMRVSSDLPPFASHAKYGYTL-EEAOSHLEEVGLKANKYNHRLTMHPGQYTOIASPRE 366
Db 294 IRFURLSSEFPFASHPVHGYKLAFASEVLAEGRAVAAELGHRLLTHPGQFTQLGSPRK 353
QY 367 VVDSAIRDLAYHDEILSRMKLNEQLNKDAVLIIHLGGTFEGCKKETLDRFRKNYQRLSDS 426
Db 354 EVVESAIRLEYHDEILSLKLPQOQNDAVMIIHMGQFGDKAATLRFKRNRYARLSQS 413

QY 427 VKARLVLENDVSWSVQDLPLCQELNIPVLVDWHHNTV--PGTLREGSLDL--PLIP 482
Db 414 CKNRVLVLENDVQVTHDILLPVCEELNIPVLDYHHNCFDPAHLREGTLDISDPKLOE 473
QY 483 TIRETWKGIITOKOHSYSESADPTAISGMKRRRAHSRVDFFPCDPDTMDLMEAKEKEQA 542
Db 474 RIANTWKRGKIGKMIHSEPCD--GAVTPRHRRKRRPRVMTLPPCPDMDLMEAKEKEQA 532
QY 543 VFELCRRYEL-----QNPCC-----LEIMGPE---- 565
Db 533 VFELMTRFKLPGPEKINDMVPYDRDDENRPPVKAAPKKKGGKRRKRTTDEAAAEVVD 592
QY 566 --YDQTRDG-----YPPGAEKRLTARKRRSRK-----EEVEED 597
Db 593 TAADDVKDAPEGKPEVEERAMGPPYVWPLGCEEWLKPKKREVKKGVPEVEDE 651
RESULT 3
S55418
UV-endonuclease homolog ywjd - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 15-Oct-1999
C:Accession: S55418; B70060
R:Glaser, P.; Danchin, A.
submitted to the EMBL Data Library, May 1995
A:Description: Cloning and sequencing of the Bacillus subtilis chromosomal region fro
A:Reference number: S55414
A:Accession: S55418
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-320 <GLA>
A:Cross-references: EMBL:Z49782; NID:g853752; PID:CAA89865.1; PID:g853757
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabbret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koettel, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A:Reference number: A69580; MUID:98044033
A:Accession: B70060
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-320 <KUN>
A:Cross-references: GB:Z99123; GB:AL009126; NID:g2636240; PID:CAB15748.1; PID:ell862
A:Experimental source: strain 168
C:Genetics:
A:Gene: ywjd

Query Match 11.6%; Score 360; DB 2; Length 320;
Best Local Similarity 29.8%; Pred. No. 6.6e-13;
Matches 84; Conservative 61; Mismatches 109; Indels 28; Gaps 8;

QY 276 TIORDGLSVKQOLGTONVLDLKLIVENHNFGIHFMRVSSDLPPFASH--AKGYTLEFA 333
Db 34 TERKEALITVTKANLRNTM---RTLHYIIGHGILPYRFSISSIVPLATHPDVWMDVETPF- 89
QY 334 OSHLEEVGLKANKYNHRLTMHPGQYTOIASPREVVDSAIRDLAYHDEILSRMKLNEQLN 393
Db 90 QKEFPEIGELVKTHQRLTSEHPNQFTLTPSKESVTKNVATDMAYHYRMLEAMGIADR-- 147
QY 394 KDAVLIIHLGGTFEGCKKETLDRFRKNYQRLSDSVKARLVLENDVSWSVQDLPLCQELN 453
Db 394 KDAVLIIHLGGTFEGCKKETLDRFRKNYQRLSDSVKARLVLENDVSWSVQDLPLCQELN 453

D_b 148 -SWINHHGATGNKDTATAQHNTIKQLPQELKERMTLENDKDVTYTTETLQVCEQED 205

Q_y 454 IPLVLDMWHNNIIVPGTLREGSLDMLPIPTIRETWTRKGITOKOHYESADAPTAISGMKR 513
 :|:||| | :|| :|::|:|||| | ||

D_b 206 VPEVFDFHFEYANP-----DDHADLNVALPRMIKTWERIGLPKVHLSSPKSEQAI----- 256

Q_y 514 RAHSDRV---FDPPPCD-----PTWDLMTAEKEGOAVFEL 546
 :|:||| | :||:|||:|||||:|:

D_b 257 RSHADVVDANFLPLLRLERFRWGTNIDFMTEAKQDKALLR.L 298

RESULT 4
E83974
hypothetical protein BH2597 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C:Accession: E83974
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.: Hirai,
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
its relationship to other members of the family Halobacteriaceae.
A;Reference number: AB3650; MUID:20263314
A:Accession: E83974
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-322 <STO>
A:Cross-references: GB:AP001516; GB:BA000004; NID:g10175192; PIDN:BAH06316.1; GSPDB:GNOC
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH2597

BIOCHIMICA ET BIOPHYSICA ACTA

Query Match 9.8%; Score 306.5; DB 2; Length 322;
Best Local Similarity 26.6%; Pred. No. 5.7e-10;
Matches 85; Conservative 75; Mismatches 123; Indels 37; Gaps 9;

QY 247 RGRLGACLTILRSMKRVFCSTCRITTIOR-----DGLESVKOLGTQNVLDLIKLVE 301
Db : || :: : : : | : || : : : : : : : : :
2 RIQFGIVANSMELANASP-----SKTWATQFPKIEDHEAGLRKKLERIAKTNLHNCURLLK 57
QY 302 WNIINFGIHPWRVSSDLPFPASH---AKYGTYLEFAOSHLEEYCKLANKYNHRLTTHPGGY 358
Db : | : | : ||| : : : : : : : : : : : : : : :
58 HNLAYQISFPRLSKKLVPLNVNPHLTEGWKEYLAIAB--ELQAQGEFASEHQMRIDPDPDH 116
QY 359 TQTASPREVVVDSATRDLAYDBEILSRMLNQKNDAVLIHLGCTFEKGKETLDLRFPK 418
Db :
117 VLNSEAKEITRRSLOTLLHYHKLLXGMEDIPHR----CVLHVGGKKRGAEGLQEFTIE 172
QY 419 NTQRUSDSVKARLVENDDVSVSDOLLPLCOELNTPLVDMWHHNINVPGLTRKSGLDIM 478
Db : : : : : : |||| : : : : : : : : : : : : : : :
173 NTA SIPKSLIMILENDDBDKSYTIDDVLVGKLALPVVLDITHHDVL-----HRSKSLQ 227
QY 479 PLIPTIREWTWRTGITOKOHYSSESADPTAISCMKRPAH-----SDRPVDF-----PPCPD 528
Db :
228 ETWORIVATWEBSPLVPVKIHL-----SPLSGEDPRHHDIYNADRFTAFLHEIGADA VD 282
QY 529 TWDLMTAEAKEQAVPELCR 548
Db : : |||| : : : : : : : : : : : : : : :
283 HLHVMIETAKKDLALFQLMK 302

RESULT 5
C75350 Probable UV damage endonuclease - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: C75350
M.: Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; PMID:20036896
A:Accession: C75350

A;Status: preliminary
A:Molecule type: DNA
A:Residues: 1-326 <WHI>
A:Cross-references: GB:AE002022; GB:AE005590; PIDN:AAF1370.1; PID:g645
A:Experimental source: strain R1
C;Genetics:
A:Gene: DR1819
A:Map position: 1

[illegible]

RESULT 6

S55785

nucleolar protein gar2 - fission yeast (Schizosaccharomyces pombe)

N:Alternate names: probable RNA-binding protein RBD18

C:Species: Schizosaccharomyces pombe

C:Date: 28-Oct-1996 #sequence_revision 07-Feb-1997 #text_change 17-Nov-2000

C:Accession: S55785; S68087; T37634; S52318

R:Gulli, M.P.; Girard, J.P.; Zabetakis, D.; Lapeyre, B.; Melese, T.; Caizergues-Ferre

Nucleic Acids Res. 23, 1912-1918, 1995

A:Title: gar2 is a nucleolar protein from Schizosaccharomyces pombe required for 18S

A:Reference number: S55785; MUID:95319932

A:Accession: S55785

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-500 <GUL>

A:Cross-references: EMBL:Z48166; NID:g663261; PIDN:CAA88179.1; PID:g663262

R:VanHoy, R.W.; Wise, J.A.

Curr. Genet. 29, 307-315, 1996

A:Title: Molecular analysis of a novel Schizosaccharomyces pombe gene containing two

A:Reference number: S68083; MUID:96171513

A:Accession: S68087

A:Status: nucleic acid sequence not shown; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 374-407 <VAN>

R:Brown, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.

submitted to the EMBL Data Library, September 1997

A:Reference number: Z21733

A:Accession: T37634

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-338, 'S', 340-500 <BRO>

A:Cross-references: EMBL:Z99091; PIDN:CAB11772.1; GSPDB:GN000066; SPDB:SPAC13F5.09

A:Experimental source: strain 972h-; cosmid cl3F5

C:Genetics:

QY 350 RL-----TMHPGOYQIASPREVVVDSAIRDLAYHDEILSRMLNQLNKLDAVLIIH 401
 Db 553 RDLKKKNETLHN-----DIANKNSYEEYKENGKLSERLN----- 590
 QY 402 LGCTPFGKKTDLDRFNKYNQRLSDSVKARLIVLENDVSVSDVLLPLCOELNPLVLWH 461
 Db 591 ---ILOQYNTLONVKSNSNEHSDIKROCELNVLKKESTKILSLEDLN----- 640
 QY 462 HHNVIPGTLREGS 474
 Db 641 YANIVQDKTREAN 653
 RESULT 9
 T34036
 hypothetical protein B0041.7 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T34036
 R:Fulton, R.; Wohldmann, P.
 A:Description: The sequence of C. elegans cosmid B0041.
 A:Reference number: 221466
 A:Accession: T34036
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1359 <FUL>
 A:Cross-references: EMBL:AF000196; PIDN:AAC24256.1; GSPDB:GN00019; CESP:B0041.7
 A:Experimental source: strain Bristol N2; clone B0041
 C:Genetics:
 A:Gene: CESP:B0041.7
 A:Map position: 1
 A:Introns: 12/2; 59/2; 248/2; 582/1; 675/1; 733/3; 966/2; 1044/2; 1310/3

Query Match 4.8%; Score 148.5; DB 2; Length 1359;
 Best Local Similarity 20.9%; Pred. No. 1.7;
 Matches 136; Conservative 96; Mismatches 222; Indels 197; Gaps 34;
 QY 83 LKNETELANISGPHKKKSTSTSTRKRARSKKKATDSVSKIDESVASYDSSTHLRSS 142
 Db 229 VKRESE---DEAPEKK--KTEKRARSKTSESESE--SEKSDDEEEKESSP----KP 277
 QY 143 RSKKPNYNYS--SSESESEQ-----ISKATKKVKQKEEYEEVEVDEKS 185
 Db 278 KKKKPLAVKKLSSDESESDVEVLPPQKKRGAVTLISDEDEKQKSESE--ASDVEEKV 336
 QY 18R LKNESSDPEFPVPPQLETPISKRRRSRSKAKLEKSTMNLD-----DHAPR 234
 Db 337 SKKAKKQESSESSESGSDSEGSITVNRKSKKKEKPKKKGIIMDSKLOKETIDAERA 396
 QY 235 EMFDCLDKPIPMWRGLGYACLNTI-----LRSMKERVFCSTCRITTIORDGL---E 283
 Db 397 ERKRREKK-----QKEFNCIVLEEGEDTEMTLTGTSSQKUKSVVLPDSDSTVDEE 448
 QY 284 SVKOLGTQNVLDLKLVEVNNHNFHIFMRVSSDLFPFASHAKYGYTLEPAQ----- 334
 Db 449 SKRPVEVHN--SLVRILKPHQAHGIQFM-----YDCAFESLDRLDTGS 490
 QY 335 -----SHLEEVGKLAN--KYNHRLTMHP--GOYTQ-----IASPREVVVD---SAIRLAYHD 380
 Db 491 GGLAHCMGIGKTLQVITELHTVLMHEKIGEKCKRVLVVVPKNVNIINWPEKQKWLVDND 550
 QY 381 EILSRMLNQLNKLDAVLIIHGGTFEGKKT-----DRFR----- 417
 Db 551 BELDTIDVNE-----LDSYKTIEDRRALKAWHSSKTPSVMIIGYDLFRILTVEED 601
 QY 418 -----KNYORLSDSVKARL-----VLENDVSVSDVLLPLCOELN 453
 Db 602 PKKKPKPNRRNRLEKAKEDFRKYLQNPQDPVMVVCDEAHKLNKDSALS--KCMVKILTKRR 660
 QY 454 I-----PLVLDWHHHNVPGTLREGSLDMLPLIPTIRETWTRKGTQKOHYESADPTAI 508
 Db 454 I-----PLVLDWHHHNVPGTLREGSLDMLPLIPTIRETWTRKGTQKOHYESADPTAI 508

Db 661 ICLTCTPLONNLMYHVMNFVKPG-----LLGKTTEFANFRVFNILNRGRTKDASPLEV 714
 QY 509 SCMKPRAHSDRVFD--FPCCDPTMD--LMIEAKEKEQAVFELCRRYELONPPCPL----- 559
 Db 715 SFMKRRCHV--LYDHLKWCYDKRKDYRVLTALPPKQYEVINVROTERO---CALYNFLN 769
 QY 560 ETMG-----PEYDQ--TRDGYVPPCA---EKRLTARKRSRKEEVEED 597
 Db 770 DIVGDSGLSKRLLPDYHMFSTRITWHPYQVLVLEQRM--ERERVMDAEAE 818
 RESULT 10
 S27833
 rhoptry-associated protein 1 precursor - malaria parasite (Plasmodium falciparum)
 N:Alternate names: protective antigen
 C:Species: Plasmodium falciparum
 C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Jun-2000
 C:Accession: A45514; S27833
 R:Ridley, R.G.; Takacs, B.; Lahm, H.W.; Delves, C.J.; Goman, M.; Matile, H
 Mol. Biochem. Parasitol. 41, 125-134, 1990
 A:Title: Characterisation and sequence of a protective rhoptry antigen from Plasmodium falciparum
 A:Reference number: A45514; MUID:90348711
 A:Accession: A45514
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-782 <R12>
 A:Cross-references: GB:M32853; NID:g160656; PID:g160657
 C:Superfamily: Plasmodium falciparum rhoptry-associated protein 1
 Query Match 4.7%; Score 147.5; DB 2; Length 782;
 Best Local Similarity 20.8%; Pred. No. 0.94;
 Matches 146; Conservative 95; Mismatches 241; Indels 219; Gaps 39;
 QY 43 FHCPLDPTLKSLPMSSKTTLSMLPOVNIIGANSFSAETPVDLKKNETELANISGPHKKST 102
 Db 122 FKASPVVKTSTVPSQTQ-----SGLKSSSPSTKSSSPNKSAS--PHGESN 168
 QY 103 STSTKRRARSKKATDSVSKIDESV--ASYDSSTHLRRSRSKPKPNYN-----SSSES 156
 Db 169 S--SESTTKSKRSASVAGIVGADEEAPAPKNTLTPL-----EELYPTNVNLFNKKYSLN 223
 QY 157 ESEGIKATKK--VKQKEEYEEVEVD-----EKSLEKNESSDEFPVVPBQ---LE 204
 Db 224 NMEENILNKGDLVAQKEFEYDENNEKAKQDKKALEKIGKQSDPEPFMFSENKFL 283
 QY 205 TPKSKRRRSRSKAKLEKSTMNLDHAPREM-----FDCDKPIPMWRG----- 248
 Db 284 NOVKERNVAGSFRFSKLNPFKKEDEVIEKTEVSKKTPSGIGFNLTKQKAVLVGATYQ 343
 QY 249 ---RLGYACLNT-----ILRSMKERV--FCSTCRITTIORDGLSVKOLGTQNVLDLI 297
 Db 344 EYPETMLYCNPNNSMLFTTIESLOGRIIDIKKRESMASTTTFEQQKECLNMG---VLDL- 399
 QY 298 KLVENNH-----NFGIHFMRV--SSDLFPFASHAKY-----GYTLE---FAQS 335
 Db 400 ---ELNDTQCKFGTCIGSGFGEHLRLYEFENDLKFHPNIDYLTADGYKLOKNNHIELS 456
 QY 336 H-----LEEVGKLAN-----KYNHR-----LTMHPOG--YTQIASP 364
 Db 457 HVNFCLLNPKTLEEFLLKKKELKDLMGDDLLIKYENFDFNFSISITCHIESLIYDIEAS 516
 QY 365 REV-----VVDLSAI-----ROLAYHDEILSRMKNL-----POLN--KDAVLII- 400
 Db 517 QDIAAVLKIASKHLVITISGLSYKARKLVY--KIYSEIQKNPDDELYEKLTIWIYNIYMIK 574
 QY 401 --HLGGTFEGKKTLDREPKN-----YORLSDSVK-----ARLVLENDVDSW---SV 442
 Db 575 RYTYAYALEGVCVYLEHDKSOMYTELHYNKIVDSVRYSSCFKNVIVYNAIISGIEHI 634
 QY 443 QDLPLCOELNPLVLDWHHHNVPGTLREGSLDMLPLIPTIRETWTRKGTQKOHYES 502
 Db 635 KHFLKLVPRHN--FLDDYHFNISIF-----EKEIKPAKKYSTS 669

QY 503 ---ADPTAIS-----CMKRAHSDRVDFPPCDPTMDLMEAKEKEQAVFELCRRYELONP 555
DB 670 HYFDPTVASAYYALNDRMTWTIINDY-----FEAKKELATIVSRMTDM--- 716
QY 556 PCPLEIMGPEYDQTRDGYYPGCAEKRLTARKRRSRKEVEE 596
DB 717 ---LSLONEESKIPND-----KSANSKLATRLMKFKAEIRD 750
RESULT 11
RN283
C:Species: parainfluenza virus type 1
C:Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 16-Jul-1999
C:Accession: E40234
R:Power, U.F.; Ryan, K.W.; Portner, A.
Virology 189, 340-343, 1992
A:Title: The P genes of human parainfluenza virus type 1 clinical isolates are polycistronic
A:Reference number: A40234; MUID:92295573
A:Accession: E40234
A:Molecule type: genomic RNA
A:Residues: 1-368 <POM>
A:Cross-references: GB:M74080; NID:g332679; PIDN:AAA46830.1; PID:g332681
C:Superfamily: parainfluenza virus polymerase-associated nucleocapsid phosphoprotein
C:Keywords: nucleocapsid; phosphoprotein

Query Match 4.7%; Score 147; DB 1; Length 568;
Best Local Similarity 21.2%; Pred. No. 0.66;
Matches 114; Conservative 87; Mismatches 214; Indels 122; Gaps 23;
QY 62 LSMLPQVIGANSFSAETPVDLKKENETELANISGPHKKSTSTTRKARRSKKATDSV 121
DB 36 LSKYP-TEIGEDRWLHNIIDNPKENKSSCKSDONNKDRAISTQDH-RSEB---SGI 90
QY 122 SKIDESVASYDSSTHLR-----RSR---SKPPVYNSSSE-----SSEE 160
DB 91 SRTGES-----KTETHARILQOQHTRASRGTSNPLPENGMGDERNTRIDEDSPNRRH 146
QY 161 QISKATKVKQ-----KEEYVEVEDEKSLKNESSEDEFEPVVPQETPIKRRRSR 214
DB 147 QRSVPTDEDRKMAENSCKREDOVEGFPPEVVGSTLSDDGE-----GRTN 192
QY 215 SSAKNLERESTMNLDDHAPREMFCDLKPIPWGRGLYACLTILRSKMERVFCSTCR 274
DB 193 NNGRSMETSST-----HSTR-ITDVTNPSP-----ELEEAVALQKKRRP 231
QY 275 TTIQDGLSVKQLGTQNVLDLILVW-NHNFGRHMRVSSDLFPFASHAKYGYTLEFA 333
DB 232 TTIKRQTRSR---TQN-SLHKSTGSDSNLEHDHNTKTSQKIPP----- 273
QY 334 QSHLEEVGKLANKYNHRLTMHPGOYTQIASPREVVVDSAIRDLAYHDEILSRMKLNQLN 393
DB 274 SKNEEPAVTKNNHNRKTKHT---TNANNNAKCLPTPEHDTTSNEBGTNTSDVEM-- 328
QY 394 KDAVLIHLG-----GTEGKETILDRFK-----NYQRLSDSVKARLVLENDVSWV 442
DB 329 --AKLLVSLGVNKSQHEFELSRRASHOPAKRLKSANYKEMTNLCGLMLLSVEKSLGNKV 386
QY 443 QDLLPLQCELNPLVLDHWHINIVPGTREGSLDMLPIPTIRETWTRKGITQKHYES 502
DB 387 EENRLLTKIOIEIDSSDLHRRFSEYQKQNSLMMANLSTLHIITDRGKT-----GDP 441
QY 503 ADPTAISGMKRAHSDRV-----FDFPPCDPTMDLMEAKEKEQAVFELCRRYELONP 555
DB 442 SDTTRSPSVFTKGKONKVKKTRF-----DPSNEALGGQEFKPDILIREDELREDIRNP 493

RESULT 12
T38786
translation initiation factor eIF-3 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Sep-2000
C:Accession: T38786; T38056
R:Skelton, J.; Churcher, C.M.; Barrrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z21751
A:Accession: T38786
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 280-913 <SKE>
A:Cross-references: EMBL:Z98762; PIDN:CAB11485.1; GSPDB:GN000066; SPDB:SPAC4A8.16c
A:Accession: T38056
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-401 <SK2>
A:Cross-references: EMBL:Z98599; PIDN:CAB11250.2; GSPDB:GN000066; SPDB:SPAC1E11.01c
A:Experimental source: strain 972h; cosmid c1E11
C:Genetics:
A:Gene: SPDB:SPAC1E11.01c
A:Map position: 1

Query Match 4.7%; Score 145.5; DB 2; Length 918;
Best Local Similarity 23.5%; Pred. No. 1.5;
Matches 103; Conservative 57; Mismatches 172; Indels 107; Gaps 20;
QY 71 GANSFSAETPVDLKKENETELANISGPHKKSTSTTRKARRSKKATDSVSKIDESVA 130
DB 7 GGSDDSDAESVDSSEEN---RLTSRLKKODSSSESESESESESESESESESE 61
QY 131 SYDSTHLRRSSKKPPVYNSSSESESEBEQISKATKKYKQKEEYVEVEDEKSLKNES 190
DB 62 SESEVEVPK---KKAASSEDSESDSE-----SSEEEETESEDSEVSESESESE 112
QY 191 SSDEFPVVPQETLPIKRRRSRSKAKNLEK-----STMNLDDHAPREM 236
DB 113 ESSESESESE-ESDESERSGSPFLKKPEKEAPAGLKLFGESSESEDEEGRV 171
QY 237 F-DCLDKPIPWGRGLYACLTILRSKMERVFCSTCRITTTIQDGLSVKQLGTQNVLD 295
DB 172 VKSAKD---RYEEFISCMETIKNAMS-----SNWIVVSNEFDHLNKVSQ----- 214
QY 296 LILVWNNHNF--IHFMRVSSDLFPFASHAKYGYTLEFAQSHLEEVGKLANKYNHRLT- 352
DB 215 --KCKEAGRNPPYIEFLSALDOKLESADKA-----FIKSLDAANGRAFALQORVK 265
QY 353 -----MHPGOYTQIASPR-----EVVWDS-AIRDLAYHDEILSRM 386
DB 266 NNFQFQSDIDRYRKDPGEFMKPAELNEIPKAGKAGQDEVIVDGVATRGIVAPTEGLGK- 324
QY 387 KLFQNLKQDAVLIHLGLGTFE--GKKETILDRFK-----NYQRLSDSVKARL 431
DB 325 --PEET-PADIEFKYLRALFEARGKKST-DRSEQIRLLEKLSIAVTDTYORL--RVKVAL 378
QY 432 VLENDVSWVQDILLPLCQ 450
DB 379 LAVREDINTSGQYMPIDQ 397

RESULT 13
S48326
hypothetical protein YML071c - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C:Date: 02-Dec-1994 #sequence_revision 09-Mar-1996 #text_change 29-Oct-1999
C:Accession: S48326; S48825
R:Bowman, S.; Churcher, C.
submitted to the EMBL Data Library, August 1994
A:Reference number: S48326
A:Accession: S48326
A:Molecule type: DNA
A:Residues: 1-113 <BOW>
A:Cross-references: EMBL:Z38114; NID:g558402; PIDN:CAA86249.1; PID:g558403; MIPS:YML0
R:Brown, D.; Bowman, S.
submitted to the EMBL Data Library, October 1994

A:Reference number: S48816
A:Accession: S48825
A:Molecule type: DNA
A:Residues: 88-607 <BRO>
A:Cross-references: EMBL:Z46373; NID:g587529; PIDN:CAA86507.1; PID:g587539; MIPS:YML0710
C:Genetics:
A:Map position: 13L

Query Match 4.7%; Score 145; DB 2; Length 607;
Best Local Similarity 21.6%; Pred. No. 0.93;
Matches 82; Conservative 73; Mismatches 160; Indels 64; Gaps 18;
1 MLRLKLNQIS--KRIVFTILKQKAFK-NHPCVPSVCTITYSRFHCPLDPL-----50
258 LVKLLSTNVSNLKLILQYLSNTPPDGKTNKSLLSV--FLANRYKFIETDASYPLDV 315
51 ----KSLPMSKTTLSLMPQ-----VNIGANSFSAET---PVDLKKENETELANISGPH 98
316 ESSNESLTEMVKKRIEVLREHVYMSLVFLKSFYDYNLEIPEPELESTVLRINGTN 375
99 KKSSTSTRKARSKKKATSDSKIDESVASYDSSTHLRRSRSSKKPVNYSSESSES 158
376 EEKETEKEKETKEEYQKQDSVANN--EEDVTENKSIDEDVQEEVOGKVE--GEDDGAERKT 433
159 EEOISKAT--KVKOKEEVEEYVEDESKLNKNESSDEFE--PVVPEOLETPISKRRRSRS 215
434 ENETENVTNKTEKAKEKEEVEVNTKDKAKEEEELINKVEVTPPEPSKIDKAE----490
216 SAKNLEKSTNLDHAPREMFCDLCKP IPWRGRGLYACLTILRSKMERVFCRSRTCRIT 275
491 -----KEEEINKVEVTPPEP-----SKKIRTSKRENKIPTNAVMLQFVDK-----C-IT 534
276 TIQRD---GLSVKOLGTQNVLDLI-----KLVEVNNHFGIHFMRVSSDLPPFASHAKYGY 328
535 YVLKDLTGLNSIK--LSDSVGLQLYVCSFRCLDLNRYNHLFLKINDTSLF-----585
329 TLEFAQSHLEEVGLANKY 347
586 TTEQLARAIDKRAELASKY 604

RESULT 14
A59294
skeletal myosin - nematode (Onchocerca volvulus)
N:Alternate names: major body wall myosin
C:Species: Onchocerca volvulus
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 08-Sep-2000
C:Accession: A59294
R:Werner, C.; Rajan, T.V.
Mol. Biochem. Parasitol. 50, 255-260, 1992
A:Title: Comparison of the body wall myosin heavy chain sequences from Onchocerca volvulus
A:Reference number: A59294; MUID:92158005
A:Accession: A59294
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1957 <WER>
A:Cross-references: GB:W74066; NID:g159892; PIDN:AAA29420.1; PID:g159893
A:Experimental source: sex female
C:Superfamily: myosin heavy chain; myosin motor domain homology
F:87-773/Domain: myosin motor domain homology <MMO>

Query Match 4.7%; Score 145; DB 2; Length 1957;
Best Local Similarity 18.1%; Pred. No. 4.3;
Matches 130; Conservative 111; Mismatches 231; Indels 248; Gaps 33;
83 LKKE-NEP-----ELANISGPHKSTSTSTRKARSKKKATSDSVSKIDES----128
1111 LKQVNTQNRITTELELEN-----BRQSRKAERAKSLQLELELGRLEDEQGAT 1164
129 ----VASYDSSTHLRRSRSSKKPVNYSSES-----SESEQISKATKKV 169

Db 1165 AAQVEVNNKKREAEAKLRRLDLEANNHNNHNLAAIRKKHNDVAELGDQIQEQVQKAKAKI 1224
QY 170 K-----OKEEEYVEEVEDEKS---LKNSSSDSEFEPPVPEOLETPISKRRR-----212
Db 1225 EKDKMAQERODLVAOIDGEAERMMNNKLAQYEMQAE--LQTKCDQONKOLQEFPTL 1283
QY 213 -SRSSAKNL-----EKSTMN:DDHAPREMF-----CLDKP IPWRGRGLYACLN- 256
Db 1284 KTRLSNENTDLCKQIEEAESQVNMATRLKAQLTSQLLEARRSLDEARDNNLAAQMKNY 1343
QY 257 -----TILSMKKEKVCPCSTRTITTTORGL--ESVKOLGTQNVLDLILKLVNNHNGIHF 310
Db 1344 QHEIOVLESMEFEI-----EGISELMKQLSRANA-----EIQWQOTRFESG 1386
QY 311 MRYSSDLTFPASHAKYGY-----TLEFAQSHLEEVGKLT-----ANKYN 348
Db 1387 LLKGDLEESKKQMIKMMELQETLDAANSKISSLEKTSRLVSDLDQADQMDVERANSTY 1446
QY 349 HRLTMHPGQVQTATSPRE-----VVVDSATRD-----AYHDEIL-----383
Db 1447 SQLEKQKGFQKVIDEMKKTKDDIATEVDNAQREARNVSTELFKLSEQDEVLETIEGLR 1506
QY 384 -----SRMKLNEOLN-----KDAV-----397
Db 1507 RENKLAQERKDLTDQLGEGRSVFEMQKILIRLEVEKDELQHALDQPDQDAEESKVL 1566
QY 398 -----LIIHLGGTFEGKKTLDKFRKNYORLSDSVKARLVLENDVYSWSVDLLPLC 449
Db 1567 RAOVEVSQRAETEKRIQKEEFENTRNHQRJESMOAS--LENE--TRSKADJMLRK 1622
QY 450 QEL-----NIPLVDHNNHIVPTLREG---SLDMLPLIPTITRETWRKGTQKQHY--499
Db 1623 KKLSDINELEIALD--HANQANAQAKNVKTYQDQMRLOQOVETEQRNRSRQYLN 1680
QY 500 -----SESADPTAISCMKRRASDRVDFPPCDPTMDLMEAKEQAVFELC---547
Db 1681 MEKATILLOSEKSEMSVANGQARARKADYD-----ANEAHTOCNELSAQAESLCGSR 1734
QY 548 RRYELONPPCPLEITMGPEYDQTRDGYPPCAEK-----RLTARKRRSRKEEVED 597
Db 1735 RKLDTE-----LLAQADLDTLNEY--KASBEERKAASDAARLAELQURKEQENSLQND 1787

RESULT 15
S38173
myosin-like protein MLP1 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YKR093w; protein YKR415
C:Species: Saccharomyces cerevisiae
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 21-Jul-2000
C:Accession: S38173; S40647; S31207
R:Baladron, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantale submitted to the Protein Sequence Database, March 1994
A:Reference number: S38158
A:Accession: S38173
A:Molecule type: DNA
A:Residues: 1-1875 <BAL>
A:Cross-references: EMBL:Z28320; NID:g486586; PID:g486587; MIPS:YKR093w
A:Experimental source: strain S288C
R:Bou, G.; Esteban, P.F.; Baladron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.; Yeast 9, 1349-1354, 1993
A:Title: The complete sequence of a 15 820 bp segment of Saccharomyces cerevisiae chr
A:Reference number: S40644; MUID:94205265
A:Accession: S40647
A:Molecule type: DNA
A:Residues: 1-1875 <BOU>
A:Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51948.1; PID:g450554
A:Experimental source: strain S288C
R:Koelling, R.; Nguyen, T.; Chen, E.Y.; Bolstein, D.
Mol. Gen. Genet. 237, 359-369, 1993
A:Title: A new yeast gene with a myosin-like heptad repeat structure.
A:Reference number: S31207; MUID:93247549
A:Accession: S31207
A:Molecule type: DNA

A:Residues: 1-300,'A',302-1875 <KOE>
A:Cross-references: EMBL:L01992; NID:g171958; PIDN:AAA34783.1; PID:g171959
C:Genetics:
A:Gene: SGD:MLP1
A:Cross-references: SGD:S0001803; MIPS:YKR095w
A:Map position: 11R

```
Query Match          4.6%  Score 144;  DB 2;  Length 1875;
Best Local Similarity 21.4%  Pred. No. 4.6;  Mismatches 164;  Indels 102;  Gaps 17;
Matches 92;  Conservative 71;

QY 88 ETELANISGPHKKSTSTSTRKRARRSKKATDSVSKIDESVASYDSSTHLRRSSRSRKKP 147
Db 1454 QNDLSNIVESMKKSPEDKIKFI---REKTOEVNEKILEAQERLNQPSNINMEIKKK- 1508
QY 148 VNYNSSSESE-----SEQISKATKKVKQKEEYVEEVEDEKSLKN 188
Db 1509 --WESHEQEVQKIREAEALKRIRLPTTEKINKIIERKKEELEKEFEKVEER-IKS 1565
QY 189 ESSSDEFEPVPEQLETPISKRRRSRSKAKNLEKESTMNLD-----HAPREMFDK 239
Db 1566 MEQSGEIDVVRKQLEAKVQEKQ-----KELENYNKKLOELKDVPHSHISDDERDK 1619
QY 240 LDKPIWRGRGLYACINTILRSMKERV-----CSRTCRITTIQRDGLSEVKOL----- 288
Db 1620 LRAETESRLREEF---NNELQAIKKKSFDEGKQQAAMKTTLLERKLAKMESOLSETKOSA 1676
QY 289 -----GTQN-VLDLIKLVENHNFGIHFMRVSSDLFPFASHAKYG-----YTLEFAQS 335
Db 1677 ESPPKSVNVQNPPLGLPRKIEENSNSFPNLLSGEKLKLNKSSSGGFNFTSPSPNK 1736
QY 336 HLEEVG-----KLANKYNHRLTMHPGQYTOIASPREVVVDSAIRDLAYHDEILS----- 384
Db 1737 HLQNDNDKRESLANKTDPPTHLEPS--FNIPASRGLISSSTLSTDTNDEELTSNNPAQK 1794
QY 385 -RMKLNQNLKDAVLIHLGGTFEGKK-----ETLDRFRKNYQRLSDSVKARLVLEN 435
Db 1795 DSSNRNVQSEEDT-----EKKEGEPVKRGEAIEEOTKSNKRPIDEV-GELKNDE 1843
QY 436 DDVSWSVQD 444
Db 1844 DDTTENINE 1852
```

Search completed: January 15, 2002, 13:52:22
Job time: 391 sec

GenCore version 4.5

Copyright (c) 1993 - 2000 Compugen Ltd. All rights reserved.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:50:21 ; Search time 134.53 seconds

Title: US-09-724-296-2

Perfect score: 4349

Sequence: 1 MTKPLTGLGKVLQVQPTR.....KRLTARKRRSRKEVEDEK 828

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL17:

- 1: sp_archaea:
- 2: sp_bacteria:
- 3: sp_fungi:
- 4: sp_human:
- 5: sp_invertebrate:
- 6: sp_mammal:
- 7: sp_mhc:
- 8: sp_organelle:
- 9: sp_phage:
- 10: sp_plant:
- 11: sp_rodent:
- 12: sp_virus:
- 13: sp_vertebrate:
- 14: sp_unclassified:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3112	71.6	599	3 Q10988	Q10988 schizosacch
2	1200.5	27.6	262	2 Q09195	Q09195 escherichia
3	1200.5	27.6	268	2 Q09194	Q09194 escherichia
4	1182	27.2	245	2 Q09193	Q09193 escherichia
5	1181	27.2	243	2 Q09192	Q09192 escherichia
6	1181	27.2	244	2 Q09191	Q09191 escherichia
7	1043	24.0	218	5 Q94745	Q94745 schistosoma
8	1011.5	23.3	656	3 Q01408	Q01408 neorospira
9	712	16.4	218	5 Q25595	Q25595 clonorchis
10	698	16.0	218	5 Q9X159	Q9X159 fasciola gi
11	510.5	11.7	218	6 Q9N0V4	Q9N0V4 bos taurus
12	496.5	11.4	223	5 Q97117	Q97117 boophilus m
13	479	11.0	218	11 Q35660	Q35660 mus musculus
14	476.5	11.0	218	6 Q9TSM5	Q9TSM5 macaca fasc
15	476	10.9	221	6 Q9BEB0	Q9BEB0 macaca fusc
16	473.5	10.9	218	6 Q9TSM4	Q9TSM4 macaca fasc
17	468.5	10.8	219	5 Q27653	Q27653 echinococcu
18	467.5	10.7	219	5 Q16058	Q16058 echinococcu
19	460.5	10.6	218	11 Q9WU21	Q9WU21 rattus norv

20	457.5	10.5	218	11 Q9DCB8	Q9DCB8 mus musculu
21	455.5	10.5	218	11 Q9DD25	Q9DD25 mus musculu
22	446.5	10.3	225	4 Q60550	Q60550 homo sapien
23	440.5	10.1	225	6 Q9BEA9	Q9BEA9 macaca fusc
24	435	10.0	232	11 Q9D5J8	Q9D5J8 mus musculu
25	434.5	10.0	225	11 Q9Z1B2	Q9Z1B2 rattus norv
26	430.5	9.9	195	4 Q05465	Q05465 homo sapien
27	407.5	9.4	188	6 Q9MZB4	Q9MZB4 capra hircu
28	378	8.7	219	5 Q9U582	Q9U582 psoroptes o
29	359	8.3	317	2 Q9L4F0	Q9L4F0 bacillus ce
30	359	8.3	325	2 Q9L4E9	Q9L4E9 bacillus ce
31	306.5	7.0	322	2 Q9K9P8	Q9K9P8 bacillus ha
32	285.5	6.6	125	6 Q29583	Q29583 sus scrofa
33	253.5	5.8	208	5 Q02636	Q02636 brugia mala
34	253	5.8	208	13 Q9DDU5	Q9DDU5 brachydanio
35	246	5.7	210	13 P81942	P81942 bufo bufo (
36	244.5	5.6	208	5 Q27711	Q27711 onchocerca
37	240	5.5	210	4 Q15690	Q15690 homo sapien
38	240	5.5	210	4 Q00460	Q00460 homo sapien
39	236	5.4	180	5 Q9NHB2	Q9NHB2 mytilus edu
40	233	5.4	209	6 Q9TTY8	Q9TTY8 capra hircu
41	229.5	5.3	305	2 Q9S0M9	Q9S0M9 deinococcus
42	229.5	5.3	326	2 Q9RTE6	Q9RTE6 deinococcus
43	213	4.9	216	13 Q9W647	Q9W647 oncorhynch
44	212.5	4.9	82	4 Q9UE37	Q9UE37 homo sapien
45	193.5	4.4	221	11 Q9DCU1	Q9DCU1 mus musculu

ALIGNMENTS

RESULT	1				
Q10988					
ID	Q10988	PRELIMINARY:	PRT:	599 AA.	
AC	Q10988	P87339:			
DT	01-NOV-1996	(TREMBLrel. 01, Created)			
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)			
DT	01-NOV-1998	(TREMBLrel. 08, Last annotation update)			
DE	UV-ENDONUCLEASE.				
GN	UVDE OR UVEL+.				
OS	Schizosaccharomyces pombe (Fission yeast).				
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;				
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;				
OC	Schizosaccharomycetes.				
OX	NCBI_TaxID=4896;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Sp972;				
RX	MEDLINE=96188860; PubMed=8614629;				
RA	Takao M., Yonemasa R., Yamamoto K., Yasui A.;				
RT	"Characterization of a UV endonuclease gene from the fission yeast				
RL	Nucleic Acids Res. 24:1267-1271(1996).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=97175806; PubMed=9023111;				
RA	Davey S., Nass M.L., Ferrer J.V., Sidik K., Eisenberger A.;				
RA	Mitchell D.L., Freyer G.A.;				
RT	"The fission yeast UVDR DNA repair pathway is inducible."				
RL	Nucleic Acids Res. 25:1002-1008(1997).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=972H-;				
RA	Lyne M., Rajandream M.A., Barrell B.G., Lucas M., Gaillardin C.;				
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.				
DR	EMBL: D78571; BAA11415.1;				
DR	EMBL: U78487; AAC49664.1;				
DR	EMBL: AL023859; CAA19577.1;				
KW	Endonuclease.				
SO	SEQUENCE 599 AA; 68816 MW; C09ED7219D5895BE CRC64;				

Query Match

71.6%; Score 3112; DB 3; Length 599;

0.6-1.0

```

Best Local Similarity 100.0%; Pred. No. 6e-190;
Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 MRLLLKNIQISKRIVETILKOKAFKGNHPCVSVCTIYSRFHCLPDLTKSLLPMSST 289
Db 1 MRLRLKNIQISKRIVETILKOKAFKGNHPCVSVCTIYSRFHCLPDLTKSLLPMSST 60
QY 290 TSLMPOVNI GANFSFAETPVDLKKENETELANISGPHKKSTSTSRKARSSKKKATDS 349
Db 61 TSLMPOVNI GANFSFAETPVDLKKENETELANISGPHKKSTSTSRKARSSKKKATDS 120
QY 350 VSDKIDSVASYSSTHLRSSRSKPPVNVSSSESESEEQISKATKKVKQKEEYVEE 409
Db 121 VSDKIDSVASYSSTHLRSSRSKPPVNVSSSESESESEEQISKATKKVKQKEEYVEE 180
QY 410 VDEKSLKNSSSDEFEPVPEQLETPISKRRRSRSKAKLEKSTWNLDDHAPREMFDC 469
Db 181 VDEKSLKNSSSDEFEPVPEQLETPISKRRRSRSKAKLEKSTWNLDDHAPREMFDC 240
QY 470 DKPIWGRGLGYACLNITLRSKRVFCSTCRITTIQRDGLSVKQLGTQNVLDLIKLV 529
Db 241 DKPIWGRGLGYACLNITLRSKRVFCSTCRITTIQRDGLSVKQLGTQNVLDLIKLV 300
QY 530 EWNHFGIHPMRYSSDLFPFASHAKYGYTLEFAOSHLEEVGKLANKYNHRLTMHPQYTO 589
Db 301 EWNHFGIHPMRYSSDLFPFASHAKYGYTLEFAOSHLEEVGKLANKYNHRLTMHPQYTO 360
QY 590 IASPREVVDSATRLAYHDEILSRMKLNEQNKDAVLIHLLGGTFEGKKETLDRFRKNY 649
Db 361 IASPREVVDSATRLAYHDEILSRMKLNEQNKDAVLIHLLGGTFEGKKETLDRFRKNY 420
QY 650 QRLSDSVKARLVENDVSWSDLLPLCOELNIPVLVDWHHNNIVPGTLREGSLDLMLPL 709
Db 421 QRLSDSVKARLVENDVSWSDLLPLCOELNIPVLVDWHHNNIVPGTLREGSLDLMLPL 480
QY 710 IPTRETWTGKTQKOHYSSESADPTAISGMKRRRAHSDRVDFPPCDPTMDLMIEAKEKE 769
Db 481 IPTRETWTGKTQKOHYSSESADPTAISGMKRRRAHSDRVDFPPCDPTMDLMIEAKEKE 540
QY 770 QAVFELCRRYELONPCPLEINGPEYDQTRDGYPPGAEKRLTARKRRSRKKEVEEDEK 828
Db 541 QAVFELCRRYELONPCPLEINGPEYDQTRDGYPPGAEKRLTARKRRSRKKEVEEDEK 599

RESULT 2
O09195
ID O09195 PRELIMINARY; PRT: 262 AA.
AC O09195;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DE GST-6HIS.
GN GST.
OS Escherichia coli.
OC Plasmid pGEX-2T-6His-PL2.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RA Cahill M.A.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: U84572; AAB41883.1; -.
DR HSP; P08515; IGNE.
DR InterPro: IPR000521; GST.
DR Pfam: PF00043; GST; 1.
SQ SEQUENCE 262 AA; 30392 MW; BA4C4593CB20226E CRC64;

Query Match 27.6%; Score 1200.5; DB 2: Length 268;
Best Local Similarity 97.8%; Pred. No. 7.4e-69;
Matches 224; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 5 PILGYWKIKGLVQPTRLLEYLEEKYEELHYERDEGDKWRNKKFELGLEFPNLPYYIDGD 64
Db 3 PILGYWKIKGLVQPTRLLEYLEEKYEELHYERDEGDKWRNKKFELGLEFPNLPYYIDGD 62
QY 65 VKLTQSMALIRYIADKHNMLGGCPKERAETSMLEGAVLDIRYGVSRITAYSKDFETLKVDF 124
Db 63 VKLTQSMALIRYIADKHNMLGGCPKERAETSMLEGAVLDIRYGVSRITAYSKDFETLKVDF 122
QY 125 LSKLPEMLKMFEDRLCHKTYLNGDHVTHPDMFLYDALDVLYMDPCLDAFPKLVCFKKR 184
Db 123 LSKLPEMLKMFEDRLCHKTYLNGDHVTHPDMFLYDALDVLYMDPCLDAFPKLVCFKKR 182
QY 185 TEATPQIDKYLKSSKYIAWPLQGQWATFGGGDHPKSD-LVPRGSPVEM 233
Db 183 TEATPQIDKYLKSSKYIAWPLQGQWATFGGGDHPKSD-LVPRGSPVEM 230

RESULT 3
O09194
ID O09194 PRELIMINARY; PRT: 268 AA.
AC O09194;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DE GST-6HIS.
GN GST.
OS Escherichia coli.
OC Plasmid pGEX-2T-6H.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RA Cahill M.A.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: U84571; AAB41882.1; -.
DR HSP; P08515; IGNE.
DR InterPro: IPR000521; GST.
DR Pfam: PF00043; GST; 1.
SQ SEQUENCE 268 AA; 31000 MW; 85820375FE60E625 CRC64;

Query Match 27.6%; Score 1200.5; DB 2: Length 268;
Best Local Similarity 97.8%; Pred. No. 7.4e-69;
Matches 224; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 5 PILGYWKIKGLVQPTRLLEYLEEKYEELHYERDEGDKWRNKKFELGLEFPNLPYYIDGD 64
Db 3 PILGYWKIKGLVQPTRLLEYLEEKYEELHYERDEGDKWRNKKFELGLEFPNLPYYIDGD 62
QY 65 VKLTQSMALIRYIADKHNMLGGCPKERAETSMLEGAVLDIRYGVSRITAYSKDFETLKVDF 124
Db 63 VKLTQSMALIRYIADKHNMLGGCPKERAETSMLEGAVLDIRYGVSRITAYSKDFETLKVDF 122
QY 125 LSKLPEMLKMFEDRLCHKTYLNGDHVTHPDMFLYDALDVLYMDPCLDAFPKLVCFKKR 184
Db 123 LSKLPEMLKMFEDRLCHKTYLNGDHVTHPDMFLYDALDVLYMDPCLDAFPKLVCFKKR 182
QY 185 TEATPQIDKYLKSSKYIAWPLQGQWATFGGGDHPKSD-LVPRGSPVEM 233
Db 183 TEATPQIDKYLKSSKYIAWPLQGQWATFGGGDHPKSD-LVPRGSPVEM 230

RESULT 4
O09193
ID O09193 PRELIMINARY; PRT: 245 AA.
AC O09193;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)

```

01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE GLUTATHIONE S-TRANSFERASE.
 GN GST.
 OS Escherichia coli.
 OG Plasmid pGEX-6p-3.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88329742; PubMed=3047011;
 RA Smith D.B., Johnson K.S.;
 RT "Single-step purification of polypeptides expressed in Escherichia
 coli as fusions with glutathione S-transferase.";
 RL Gene 67:31-40(1988).
 RN [2]
 RP SEQUENCE OF 221-227 FROM N.A.
 RX MEDLINE=90264387; PubMed=2160953;
 RA Cordingley M.G., Callahan P.L., Sardana V.V., Garsky V.M.,
 Colono R.J.;
 RT "Substrate requirements of human rhinovirus 3C protease for peptide
 cleavage in vitro.";
 RL J. Biol. Chem. 265:9062-9065(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Dunst R.W., Bell P.A., English J.D.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U78874; AAB37352.1; -;
 DR HSSP: P08515; IGNE.
 DR InterPro: IPR000521; GST.
 DR Pfam: PF00043; GST; 1.
 KW Transferase; Plasmid.
 SQ SEQUENCE 245 AA; 28327 MW; D51470E8A48C6CC2 CRC64;

Query Match 27.2%; Score 1182; DB 2; Length 245;
 Best Local Similarity 92.2%; Pred. No. 9.8e-68;
 Matches 226; Conservative 1; Mismatches 10; Indels 8; Gaps 2;

QY 5 PILGYWKIKGLVQPTRLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGD 64
 DB 3 PILGYWKIKGLVQPTRLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGD 62

QY 65 VKLTQSMALIRYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRISYKDFETLKVD 124
 DB 63 VKLTQSMALIRYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRISYKDFETLKVD 122

QY 125 LSKLPEMLKMFEDRLCHTKTYLNGDHVTHPDFMLYDALDVVLYMDPCLDAFPKLVCFKKR 184
 DB 123 LSKLPEMLKMFEDRLCHTKTYLNGDHVTHPDFMLYDALDVVLYMDPCLDAFPKLVCFKKR 182

QY 185 IEAIPQIDKYLKSSKYIAWPLOGWOATFGGDHPPKSDHLV----PRGSMRLRLKRNIOI 240
 DB 183 IEAIPQIDKYLKSSKYIAWPLOGWOATFGGDHPPKSDHLV----NSRVDS 238

QY 241 SKRIV 245
 DB 239 SGRIV 243

RESULT 5
 ID 009192 PRELIMINARY; PRT: 243 AA.
 AC 009192;
 DT 01-JUL-1997 (TReMBLrel. 04, Created)
 DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE GLUTATHIONE S-TRANSFERASE.
 GN GST.
 OS Escherichia coli.
 OG Plasmid pGEX-6p-2.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.

OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88329742; PubMed=3047011;
 RA Smith D.B., Johnson K.S.;
 RT "Single-step purification of polypeptides expressed in Escherichia
 coli as fusions with glutathione S-transferase.";
 RL Gene 67:31-40(1988).
 RN [2]
 RP SEQUENCE OF 221-227 FROM N.A.
 RX MEDLINE=90264387; PubMed=2160953;
 RA Cordingley M.G., Callahan P.L., Sardana V.V., Garsky V.M.,
 Colono R.J.;
 RT "Substrate requirements of human rhinovirus 3C protease for peptide
 cleavage in vitro.";
 RL J. Biol. Chem. 265:9062-9065(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Dunst R.W., Bell P.A., English J.D.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U78873; AAB37349.1; -;
 DR HSSP: P08515; IGNE.
 DR InterPro: IPR000521; GST.
 DR Pfam: PF00043; GST; 1.
 KW Transferase; Plasmid.
 SQ SEQUENCE 243 AA; 27908 MW; 0003C434CD76C3A6 CRC64;

Query Match 27.2%; Score 1181; DB 2; Length 243;
 Best Local Similarity 96.9%; Pred. No. 1.1e-67;
 Matches 222; Conservative 0; Mismatches 3; Indels 4; Gaps 1;

QY 5 PILGYWKIKGLVQPTRLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGD 64
 DB 3 PILGYWKIKGLVQPTRLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGD 62

QY 65 VKLTQSMALIRYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRISYKDFETLKVD 124
 DB 63 VKLTQSMALIRYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRISYKDFETLKVD 122

QY 125 LSKLPEMLKMFEDRLCHTKTYLNGDHVTHPDFMLYDALDVVLYMDPCLDAFPKLVCFKKR 184
 DB 123 LSKLPEMLKMFEDRLCHTKTYLNGDHVTHPDFMLYDALDVVLYMDPCLDAFPKLVCFKKR 182

QY 185 IEAIPQIDKYLKSSKYIAWPLOGWOATFGGDHPPKSDHLV----PRGS 229
 DB 183 IEAIPQIDKYLKSSKYIAWPLOGWOATFGGDHPPKSDHLV----PRGS 231

RESULT 6
 ID 009191 PRELIMINARY; PRT: 244 AA.
 AC 009191;
 DT 01-JUL-1997 (TReMBLrel. 04, Created)
 DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE GLUTATHIONE S-TRANSFERASE.
 GN GST.
 OS Escherichia coli.
 OG Plasmid pGEX-6p-1.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88329742; PubMed=3047011;
 RA Smith D.B., Johnson K.S.;
 RT "Single-step purification of polypeptides expressed in Escherichia
 coli as fusions with glutathione S-transferase.";
 RL Gene 67:31-40(1988).
 RN [2]
 RP SEQUENCE OF 221-227 FROM N.A.
 RX MEDLINE=90264387; PubMed=2160953;

```

RA Cordingley M.G., Callahan P.L., Sardana V.V., Garsky V.M.,
RA Colomo R.J.;
RT "Substrate requirements of human rhinovirus 3C protease for peptide
RT cleavage in vitro.";
RL J. Biol. Chem. 265:9062-9065(1990).
RN [3]
RP SEQUENCE FROM N.A.
RA Dunst R.W., Bell P.A., English J.D.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U78872; AAB37346.1; -.
DR HSSP; P08515; IGNE.
DR InterPro; IPR000521; GST.
DR Pfam; PF00043; GST; 1.
KW Transferrase; Plasmid.
SQ SEQUENCE 244 AA; 28430 MW; 7461B31F17027042 CRC64;

Query Match 27.2%; Score 1181; DB 2; Length 244;
Best Local Similarity 96.9%; Pred. No. 1.1e-67;
Matches 222; Conservative 0; Mismatches 3; Indels 4; Gaps 1;

QY 5 PILGYWKIKGLVQPTRLLEYLEEKEEHLHYERDEGDKWRNKKFELGFPNLPYYIDGD 64
DB 3 PILGYWKIKGLVQPTRLLEYLEEKEEHLHYERDEGDKWRNKKFELGFPNLPYYIDGD 62

QY 65 VKLTQSMATIRYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSR IAYSKDFETLKVDF 124
DB 63 VKLTQSMATIRYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSR IAYSKDFETLKVDF 122

QY 125 LSKLPEMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVLVYMDPCLDAFPKLVCFKKR 184
DB 123 LSKLPEMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVLVYMDPCLDAFPKLVCFKKR 182

QY 185 IEATPQIDKYLKSSKYIAWPLOGWQATFGGDDHPPK 220
DB 183 IEATPQIDKYLKSSKYIAWPLOGWQATFGGDDHPPK 218

RESULT 7
Q94745 PRELIMINARY; PRT; 218 AA.
ID O94745
AC O94745;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DE 26 KDA GLUTATHIONE S-TRANSFERASE.
OS Schistosoma mekongi.
OC Eukaryota; Metazoa; Platyhelminthes; Rhabditophora; Neodermata;
OC Trematoda; Digenea; Strigeida; Schistosomatidae; Schistosomatidae;
OC Schistosoma.
NCBI_TaxID=38744;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=THAILAND;
RX MEDLINE=98222395; PubMed=9561610;
RA Grams S.V., Grams R., Korge G., Vivanant V., Upatham S.;
RT "Cloning and sequence analysis of the 26 kDa glutathione-transferase
RT gene of Schistosoma mekongi.";
RL Southeast Asian J. Trop. Med. Public Health 28:570-574 (1997).
DR EMBL; Y07663; CAA68944.1; -.
DR HSSP; P08515; IGTA.
DR InterPro; IPR000521; GST.
DR Pfam; PF00043; GST; 1.
KW Transferrase.
SQ SEQUENCE 218 AA; 25486 MW; D3080620B19DE23A CRC64;

Query Match 24.0%; Score 1043; DB 5; Length 218;
Best Local Similarity 89.8%; Pred. No. 5.8e-59;
Matches 194; Conservative 10; Mismatches 12; Indels 0; Gaps 0;

QY 5 PILGYWKIKGLVQPTRLLEYLEEKEEHLHYERDEGDKWRNKKFELGFPNLPYYIDGD 64
DB 3 PILGYWKIKGLVQPTRLLEYLEEKEEHLHYERDEGDKWRNKKFELGFPNLPYYIDGD 62

```

```

DB 3 PILGYWKIKGLVQPTRLLEYLEEVEERLYERNEGDAWNEKFKLGLEFPNLPYYIDGD 62
QY 65 VKLTQSMATIRYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSR IAYSKDFETLKVDF 124
DB 63 VKLTQSMATIRYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSR IAYSKDFETLKVDF 122
QY 125 LSKLPEMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVLVYMDPCLDAFPKLVCFKKR 184
DB 123 LSKLPEMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVLVYMDPCLDAFPKLVCFKKR 182
QY 185 IEATPQIDKYLKSSKYIAWPLOGWQATFGGDDHPPK 220
DB 183 IEATPQIDKYLKSSKYIAWPLOGWQATFGGDDHPPK 218

RESULT 8
Q01408 PRELIMINARY; PRT; 656 AA.
ID Q01408
AC Q01408;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE UV-ENDORNUCLEASE.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=74-OR23-1A;
RX MEDLINE=95292980; PubMed=7774597;
RA Yajima H., Takao M., Yasuhira S., Zhao J.H., Ishii C., Inoue H.,
RA Yasui A.;
RT "A eukaryotic gene encoding an endonuclease that specifically repairs
RT DNA damaged by ultraviolet light.";
RL EMBL J. 14:2393-2399(1995).
DR EMBL; D11392; CAA20113.1; -.
KW Endonuclease.
SQ SEQUENCE 656 AA; 74413 MW; 4B231E7A8735C86F CRC64;

Query Match 23.3%; Score 1011.5; DB 3; Length 656;
Best Local Similarity 36.6%; Pred. No. 2.6e-56;
Matches 241; Conservative 88; Mismatches 149; Indels 181; Gaps 18;

QY 339 ARSKKKATDSVSKIDESVASYDSST-----HLRRSSR-----SKKPVNYSSESE- 386
DB 3 SRKSAALADTPQSESTFSSTLDSAPSARNLRSGRNTLQPSSEKDRDHEKRSGEEL 62
QY 387 -----SEEQISKATKKVKKEEVEEVEVDEK 413
DB 63 AGRMMCKDIANGHCLREGKEQEGVKMATEGLARMERLQRAKTKKKOLEED----- 114
QY 414 SLKNESSEDEFVPEQLETPISKRRS-----RSSAKNLEKESMTNLDHA----- 461
DB 115 GIPVPSVVSRE-PTAPYHHKSTNABEAKEPVLKTHSKDVEREAEIGVDVVKMEPAAT 173
QY 462 -----PREMFDCLDK-----PWPGRGLGYACLTNLTILSRMKERVFCRTCRIT 504
DB 174 NTEPEDAQAADAAERGAARPPAVNSSYLPWPGRGLGYACLTNLTILSRMKERVFCRTCRIT 233
QY 505 TT-----ORDGLIESVKQIGTQNVLDLKLKLVNHNFG 536
DB 234 SIVDHRHPIQFEDEPEHHLKNPKDKSKPEQDFLGHKFVQELGLANARDIVKMLCWNKYG 293
QY 537 IHFMVRSDDLPPFASHAKYGYTL-EFAOSHLEEVGKLANKYNHRLTMIPGQYTOIASPRE 595
DB 294 IREFLSSEMPFPFASHVGHYKLAPPFASEVLAEGRAAEAGLHRLTTHPGQFTOLGSPRK 353
QY 596 VVVDLSAIPLAYHDEILSRMKLNEQLNKNDAVLI THLGQTFECCKKTETLDRFRKNTQRLSDS 655
DB 354 EVVESAIRDLEHYDELLELSLLKLPQEQNDVAMI IHMGQFGDGAATLERFRKNTQRLSDS 413

```

Qy	656	VKARLVLENDVSNVODLLPQCELNIPVLVDHHTIV	--PGTLREGSLDLM--	PLIP	711
		: : : : :			
Db	414	CKNRLVLENDVGVTHDLLPVCEELINPMVLDY	HHNNICFDPAHLREGTLDISDPKQEQ	473	
		: : : : :			
Qy	712	TIRETWKRGITOKOHYESADPTALSGMKRRAHS	DRVDFPCCDPTMDLMEAKKEQA	771	
		: : : : :			
Db	474	RIANTWKRGIKOKMHPSECD--GAVTPDRRKR	PRVMTLPCCPDMDLMEAKDKQA	532	
		: : : : :			
Qy	772	VFELCRRYEL-----QNPCCP-----	LEIMGE-----	794	
		: : : : :			
Db	533	VFELMRTEFKLPGEKINDMVPYDRDDENRAP	VPVAPKKGKGRKRTTDEAAPEEVD	592	
		: : : : :			
Qy	795	--YDQPRDG-----	YTPGAEKRLTAKRRSRK-----	EVEEED	826
		: : : : :			
Db	593	TAADDVKDADEGPKVPEERAMGPNRYNVP	PLGCEEWLKPKEVKGKGVPEEVE	651	
		: : : : :			

RESULT	9
Q25595	
ID	PRELIMINARY; PRT; 218 AA.
AC	Q25595;
DT	01-NOV-1996 (TrEMBLrel_01, Created)
DT	01-NOV-1999 (TrEMBLrel_12, Last sequence update)
DE	01-JUN-2001 (TrEMBLrel_17, Last annotation update)
DE	PUTATIVE GLUTATHIONE TRANSFERASE.
OS	Clonorchis sinensis.
OC	Eukaryota; Metazoa; Platyhelminthes; Rhabditophora; Neodermata;
OC	Trematoda; Digenea; Opisthorchiida; Opisthorchioidea; Opisthorchoidea;
OC	Opisthorchiidae; Clonorchis.
OX	NCBI_Taxid=79923;
RN	{1}
RN	SEQUENCE FROM N.A.
RP	Hong S.-J.; Lee D.-H.;
RT	"Cloning and over expression of 26 kDa glutathione S-transferase from
RT	Clonorchis sinensis.";
RL	This (1996), Parasitology, Chung-Ang University College of Medicine
RL	Seoul.
RN	{2}
RN	SEQUENCE FROM N.A.
RA	Hong S.-J.;
RA	Submitted (APR-1999) to the EMBL/GenBank/DBDJ databases.
DR	EMBL; L47992; AAB46369.3; .
DR	HSSP; P31670; 1FHE.
DR	InterPro; IPR000521; GST.
DR	Pfam; PF00043; GST; 1.
KW	transferase.
SQ	SEQUENCE 218 AA; 25038 MW; 7CB17C7B837A0B7C CRC64:

[illegible]

RESULT	10
Q9XYL9	

ID	Q9XYL9	PRELIMINARY;	PRT;	218	AA.
AC	Q9XYL9;				
DT	01-NOV-1999	(TrEMBLrel. 12, Created)			
DT	01-NOV-1999	(TrEMBLrel. 12, Last sequence update)			
DT	01-JUN-2001	(TrEMBLrel. 17, Last annotation update)			
DE	GLUTATHIONE S-TRANSFERASE.				
GN	GST-1.				
OS	Fasciola gigantica.				
OC	Eukaryota; Metazoa; Platyhelminthes; Rhabditophora; Neodermata;				
OC	Trematoda; Digenea; Echinostomida; Echinostomata; Fascioloidae;				
OC	Fasciolidae; Fasciola.				
OX	NCBI_TaxID=46835;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=THAILAND;				
RA	Grams S.V., Grams R., Sobhon P., Viyanant V., Upatham E.S.;				
RT	"Molecular cloning of expressed antigens from Fasciola gigantica.";				
RL	Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF112567; AAD23997.1;				
DR	HSSP; P31670; 1PHE.				
DR	InterPro; IPR000521; GST.				
DR	Pfam; PF00043; GST; 1.				
KW	Transferase.				
SC	SEQUENCE	218	AA;	25337	MW; E9423D75C3F2EEAF CRC64;
Query Match 16.0%; Score 698; DB 5; Length 218;					
Best Local Similarity 58.7%; Pred. No. 5.1e-37;					
Matches 125; Conservative 35; Mismatches 53; Indels 0; Gaps					
Qy	7	LGYYKIKGLVQPTRLLEYLEEYEEHLYERDEGDKWRNKKFELGLFFPNLPYYIDGDVK	66		
Db	5	LGYYKIRGLQPPVRLLEYLDEEYEEHLYGRDDREKWLGDKNMGLDLPNLPYYIDDKCK	64		
Qy	67	LTQSMAIRIYADKHNMLGCCPKERAEISMLEGAVLDIRYGVSRVIAYSKDFETLKVDFLS	126		
Db	65	LTQSVAMIRYADKHGMLGSTPEERARVSIIEGAAMDRLRMGFVRVCYNPNFEEVKGDYLK	124		
Qy	127	KLPKLMKFEDRLCHKTYLNGDHWTHPDFMLYDALDVLVYMDPMLDAPFKLVCFKKRIE	186		
Db	125	ELPRTKRWSDFLGDRQYLTGSSVSHVDFVYALDCIRYLAQCLNDFPKLKEFKSRIE	184		
Qy	187	AIPDIDYKLSXKYIAMPLOGWATFGGGDHP	219		
Db	185	DLPIKAYMESEKFIKWPLNSWTASFGGGAAP	217		
RESULT	11				
Q9NOV4					
ID	Q9NOV4	PRELIMINARY;	PRT;	218	AA.
AC	Q9NOV4;				
DT	01-OCT-2000	(TrEMBLrel. 15, Created)			
DT	01-OCT-2000	(TrEMBLrel. 15, Last sequence update)			
DT	01-JUN-2001	(TrEMBLrel. 17, Last annotation update)			
DE	CLASS MU GLUTATHIONE S-TRANSFERASE.				
GN	GSTM.				
OS	Bos taurus (Bovine).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;				
OC	Bovidae; Bovinae; Bos.				
OX	NCBI_TaxID=9913;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=EYE LENS;				
RA	Jimenez-Asensio J.V., Gariand D.;				
RT	"A lens glutathione S-transferase, class mu, with thiol-specific				
RL	antioxidant activity.";				
RT	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF249588; AAF64308.1;				
DR	InterPro; IPR000521; GST.				
DR	Pfam; IPR003081; GST_mu.				
DR	Pfam; PF00043; GST; 1.				
DR	PRINTS; PR01267; GSTRNSFRASEM.				

```

KW Transferase.
SQ SEQUENCE 218 AA; 25635 MW; 3D02EA0F43C07B0A CRC64;

Query Match 11.7%; Score 510.5; DB 6; Length 218;
Best Local Similarity 47.8%; Pred. No. 4.2e-25;
Matches 98; Conservative 35; Mismatches 67; Indels 5; Gaps 1;

QY 6 ILGYWKIKGLVOPTRLLLEYLEEKYEHLVERDEG-----DKWRNKKFELGLEFPNLY 60
DB 4 ILGYWDIRGLHAIRLLLEYTDTNTEERQYSVGDADYDRSQWLNEKFKLGLDFPNLYL 63

QY 61 IDGDVKLTQSMATIRYADKHNMLGCCPKERAETISMLEGAVLDIRYGVSRAYSXDFETL 120
DB 64 IDGTHKLTQSNALTRYAKHNLCGGTEEMTRVDILENOVMDVRLAWARICYSDFPEKL 123

QY 121 KVDPLSKPLMELKMFEDRLCHTKTYLNGDVRVTHPDMFLYDALDVLYMDPCLDAFPKLV 180
DB 124 KPGFLAIEPKIKFSEFLGKRPWAGDKLTVYDVLVDLMHRIFFPKLUDAFPNLKD 183

QY 181 FKRIEAIPOIDKYLKSSKYIAWPL 205
DB 184 FISRFELKKISAYMKSSRFLPGPL 208

RESULT 12
ID 097117 PRELIMINARY: PRT; 223 AA.
AC 097117;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE GLUTATHIONE S-TRANSFERASE.
OS Boophilus microplus (Cattle tick).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Parasitiformes; Ixodida; Ixodidae; Ixodidae; Boophilus.
OX NCBI_TaxID=6941;
RN [1]
RP SEQUENCE FROM N.A.
RA He H., Chen A.C.;
RT "Characterization and molecular cloning of a glutathione S-transferase
RT from larvae of the cattle tick Boophilus microplus (Acari;
RT Ixodidae).";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF077609; AAD15991.1; -.
DR HSSP; P20136; IGSU.
DR InterPro; IPR000521; GST.
DR InterPro; IPR003081; GST_mu.
DR Pfam; PF000043; GST; 1.
DR PRINTS; PR01267; GSTRNSFRASEM.
KW Transferase.
SQ SEQUENCE 223 AA; 25604 MW; F276ACE2D8259174 CRC64;

Query Match 11.4%; Score 496.5; DB 5; Length 223;
Best Local Similarity 47.2%; Pred. No. 3.4e-24;
Matches 101; Conservative 28; Mismatches 80; Indels 5; Gaps 1;

QY 5 PILGYWKIKGLVOPTRLLLEYLEEKYEHLVERD-----EGDKWRNKKFELGLEFPNLY 59
DB 3 PVLYGWDIRGLAQPIRELLAHVDAKVDKRYTCGPPDFDRSSWLNEKTKLGLFEPNLY 62

QY 60 YIDGDVKLTQSMATIRYADKHNMLGCCPKERAETISMLEGAVLDIRYGVSRAYSXDFET 119
DB 63 YIDGDVKLTQSMATIRYADKHNMLGCCPKERAETISMLEGAVLDIRYGVSRAYSXDFET 122

QY 120 LKVDPLSKPLMELKMFEDRLCHTKTYLNGDVRVTHPDMFLYDALDVLYMDPCLDAFPKLV 179
DB 123 LKGDYLNKLPASLKAFSDYLGTHKFFAGDNLTVYDFTAYEMLAQHILFADCLKDFANLK 182

QY 180 CFKRIEAIPOIDKYLKSSKYIAWPLQWQATFG 213
DB 183 AFVDRIALPHVAAYLKSDCKIKRPLNGDMASFG 216

```

```

RESULT 13
ID 035660 PRELIMINARY: PRT; 218 AA.
AC 035660; 035661;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE GLUTATHIONE S-TRANSFERASE 6 (EC 2.5.1.18) (GLUTATHIONE-S-TRANSFERASE
DE CLASS M5).
GN GSTM6 OR GSTM5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC STRAIN=FVB, 129/SV, AND C57BL/6; TISSUE=LIVER;
RX MEDLINE=98149729; PubMed=9480867;
RA De Bruin W.C.C., te Morsche R.H.M., Wagenmans M.J.M., Alferink J.C.,
RA Townsend A.J., Wieringa B., Peters W.H.M.;
RT "Identification of a novel murine glutathione S-transferase class mu
RT gene.";
RL Biochem. J. 330:623-626(1998).
CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER OF
CC EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES (BY
CC SIMILARITY).
CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE = HX + R-S-G.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN LIVER, STOMACH AND SMALL
CC INTESTINE. NOT EXPRESSED IN SPLEEN, KIDNEY, COLON, HEART, MUSCLE,
CC BRAIN OR LUNG.
CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY.
DR EMBL; AJ000413; CAAG4061.1; -.
DR EMBL; AJ000412; CAAG4060.1; -.
DR HSSP; P09488; IGTU.
DR MGD; MGI:1309467; Gstm6.
DR InterPro; IPR000521; GST.
DR InterPro; IPR003081; GST_mu.
DR Pfam; PF000043; GST; 1.
DR PRINTS; PR01267; GSTRNSFRASEM.
KW Transferase; Multigene family.
SQ SEQUENCE 218 AA; 25628 MW; 324B20CC6C35594B CRC64;

Query Match 11.0%; Score 479; DB 11; Length 218;
Best Local Similarity 44.0%; Pred. No. 4.3e-23;
Matches 95; Conservative 37; Mismatches 78; Indels 6; Gaps 2;

QY 4 LPI-IGYWKIKGLVOPTRLLLEYLEEKYEHLVERDEG-----DKWRNKKFELGLEFPNL 57
DB 1 MPVTLGYWDIRGLHAIRLLLEYTDTNTEERQYSVGDADYDRSQWLNEKFKLGLDFPNL 60

QY 58 PYIDGDVKLTQSMATIRYADKHNMLGCCPKERAETISMLEGAVLDIRYGVSRAYSXDFE 117
DB 61 PVLIDCSHKVYQSNAILRYLGRKHNLCGTEETEEIRVVDILENVRVMDTRIONGMLCYADF 120

QY 118 ETLKVDPLSKPLMELKMFEDRLCHTKTYLNGDVRVTHPDMFLYDALDVLYMDPCLDAFPK 177
DB 121 EKKRKEFLKGLPDQKLKLYSEFLGKQWFGDKITFADFLVYDVLDOHRMFEPTCLDAFPN 180

QY 178 LVCFKRIEAIPOIDKYLKSSKYIAWPLQWQATFG 213
DB 181 LKDFMARFELKKISAYMKTSRFLSPYLKQATWG 216

RESULT 14
ID 097SM5 PRELIMINARY: PRT; 218 AA.
AC 097SM5;
DT 01-MAY-2000 (TREMBlrel. 13, Created)

```

01-MAY-2000 (Tremblrel. 13, Last sequence update)
01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE GLUTATHIONE S-TRANSFERASE MU-CLASS SUBUNIT M1 (EC 2.5.1.18).
GN GSTM1.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Wang C., Eaton D.L.;
RT "Mu-class GSTs Are Responsible for Aflatoxin B1-8,9-epoxide Activity
in the Non-human Primate Macaca fascicularis Liver.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF200709; AAF08539.1;
DR HSP: P09488; LGTU
DR InterPro: IPR000521; GST.
DR InterPro: IPR003081; GST_mu.
DR Pfam: PF00043; GST; 1.
DR PRINTS: PR01267; GSTRNSFRASEM.
KW Transferase.
SQ SEQUENCE 218 AA; 25577 MW; EE5BAD80F60C95EB CRC64;

Query Match 11.0%; Score 476.5; DB 6; Length 218;
Best Local Similarity 43.9%; Pred. No. 6.2e-23;
Matches 93; Conservative 36; Mismatches 78; Indels 5; Gaps 1;

Qy 7 LGYWKIKGLVQPTRLLEYLEEKYEHLERDEG-----DKWRNKKFELGLF 61
Db 5 LGYWDIRGLAHAIIRLLEYTSSYEKKYTMGDAPDYDRSQWLNEKFKLGLDF 64
Qy 62 DGDVKLTQSMALIRYIADKHNMGLGGCPKERAETSMLEGAVIDIRYGVSRITAYSKDFETLK 121
Db 65 DGTHTKITSNAILRYIARKHNLGCTEETEEKIRVDILENQAMVSNQLARVCYSPDFEKLK 124
Qy 122 VDFLSKLPEMLKMFEDRLCHTKYLNGLDGHVTHPDFMLYDALDVLVYMDPMLCLDAFPKLVCF 181
Db 125 PEYLEGLPTMQHFSQFLGKRPFVGDKITTFVDFLAYDLHLRIFPKLDAFPNLKDF 184
Qy 182 KRTAIPQIDKYLKSKYIAWPLQGWQATFG 213
Db 185 ISHFEGLKISAYMKSSRFLPKPLYTRVAVWG 216

RESULT 15
Q9BEB0 PRELIMINARY; PRT; 221 AA.
AC Q9BEB0;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE GLUTATHIONE TRANSFERASE M2.
OS Macaca fuscata (Japanese macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9542;
RN [1]
RP SEQUENCE FROM N.A.
RA Beuckmann C.T., Fujimori K., Urade Y.;
RL "Macaca fuscata glutathione transferase M2.";
RT Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB025799; BAB40442.1;
KW Transferase.
SQ SEQUENCE 221 AA; 26068 MW; 01F22BAC6DA57FE9 CRC64;

Query Match 10.9%; Score 476; DB 6; Length 221;
Best Local Similarity 44.3%; Pred. No. 6.8e-23;
Matches 97; Conservative 36; Mismatches 80; Indels 6; Gaps 2;

QY 1 MTKLPI-LGYWKIKGLVQPTRLLEYLEEKYEHLERDEG-----DKWRNKKFELGLF 54
Db 1 MPMPMTLGYWNIRGLAHSIRLLEYTSSYEKKYTMGDAPDYDRSQWLNEKFKLGLDF 60
QY 55 PNLPPYIDGDKLTQSMALIRYIADKHNMGLGGCPKERAETSMLEGAVIDIRYGVSRITAY 114
Db 61 PNLPLYLDGTHKITQSNAILRYIARKHNLGCTEETEEKIRVDILENQAMVSNQLARCLCYD 120
QY 115 KDFETLKVDPLSKLPEMLKMFEDRLCHTKYLNGLDGHVTHPDFMLYDALDVLVYMDPMLCLDA 174
Db 121 PDPEKLVCFKRIEAIPOIDKYLKSKYIAWPLQGWQATFG 213
QY 175 FPKLVCFKRIEAIPOIDKYLKSKYIAWPLQGWQATFG 213
Db 181 FPNLKDFISRFEGLEKISAYMKSSRFLPKPLYTRVAVWG 219

Search completed: January 15, 2002, 14:06:09
Job time: 948 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 14:06:31 ; Search time 134.53 Seconds
(without alignments)
319.662 Million cell updates/sec

Title: US-09-724-296-39

Perfect score: 1527
Sequence: 1 QLGLVCLTVGPEVRFTVTLL.....VEAKGKEAIAALRLMAPFK 294

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTRMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_Organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_prodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1520	99.5	305	2 Q9S0M9	Q9S0M9 deinococcus
2	1520	99.5	326	2 Q9RTE6	Q9RTE6 deinococcus
3	297.5	19.5	317	2 Q9L4F0	Q9L4F0 bacillus ce
4	297.5	19.5	325	2 Q9L4E9	Q9L4E9 bacillus ce
5	262.5	17.2	656	3 Q01408	Q01408 neurospora
6	250	16.4	322	2 Q9K9P8	Q9K9P8 bacillus ha
7	230.5	15.1	599	3 Q10988	Q10988 schizosacch
8	111	7.3	315	2 Q9AMN9	Q9AMN9 pseudomonas
9	102.5	6.7	437	2 Q9PFJ8	Q9PFJ8 xylella fas
10	98.5	6.5	313	1 Q9VAQ7	Q9VAQ7 aeropyrum p
11	98.5	6.5	1486	2 Q9CG24	Q9CG24 lactococcus
12	96.5	6.3	2216	10 Q9LIV03	Q9LIV03 arabidopsis
13	96	6.3	526	2 Q9X5N3	Q9X5N3 myxococcus
14	96	6.3	535	2 Q9GCB8	Q9GCB8 lactococcus
15	96	6.3	791	10 Q90962	Q90962 arabidopsis
16	95	6.2	368	2 Q9XAC4	Q9XAC4 streptomyce
17	95	6.2	515	2 Q9RVP5	Q9RVP5 deinococcus
18	95	6.2	558	11 Q9JL61	Q9JL61 mus musculu
19	94.5	6.2	388	2 Q9KID6	Q9KID6 streptomyce

20	94.5	6.2	1203	2 Q9ACX5	Q9acx5 streptomyce
21	94.5	6.2	1809	2 Q85740	Q85740 pseudomonas
22	94.5	6.2	1809	2 Q9HWG4	Q9hwg4 pseudomonas
23	94.5	6.2	2731	2 Q923T9	Q923t9 pseudomonas
24	94	6.2	898	2 Q9RI40	Q9ri40 streptomyce
25	93.5	6.1	1809	2 Q9RFM7	Q9rfm7 pseudomonas
26	93	6.1	603	2 Q9RTX0	Q9rtx0 deinococcus
27	92.5	6.1	915	11 Q9ERV7	Q9erv7 mus musculu
28	92	6.0	332	2 Q9ZGA5	Q9zga5 streptomyce
29	92	6.0	777	2 Q9AAW1	Q9aaw1 caulobacter
30	92	6.0	951	2 Q9HTC4	Q9htc4 pseudomonas
31	92	6.0	1017	4 Q9V2M0	Q9v2m0 homo sapien
32	91.5	6.0	6315	2 Q9ADL6	Q9adl6 polyangium
33	91	6.0	344	2 Q9I5Q9	Q9i5q9 pseudomonas
34	91	6.0	750	2 Q9FC09	Q9fc09 streptomyce
35	90.5	5.9	586	2 P94134	P94134 alcaligenes
36	90.5	5.9	644	10 Q9ARL6	Q9arl6 hordeum vul
37	90.5	5.9	842	2 Q9HXN5	Q9hxn5 pseudomonas
38	90.5	5.9	2040	5 Q9V696	Q9v696 drosophila
39	90.5	5.9	4247	2 Q9L8H4	Q9l8h4 streptomyce
40	90.5	5.9	11096	2 Q9L4W3	Q9l4w3 streptomyce
41	90	5.9	1421	2 Q9L8C9	Q9l8c9 polyangium
42	90	5.9	3413	2 Q54593	Q54593 amycolatops
43	89.5	5.9	319	2 Q9H295	Q9hz95 pseudomonas
44	89.5	5.9	390	2 Q9KCE3	Q9kce3 bacillus ha
45	89.5	5.9	397	2 Q9S2R3	Q9s2r3 streptomyce

ALIGNMENTS

RESULT 1

Q9S0M9 PRELIMINARY: PRT: 305 AA.

AC Q9S0M9;
DT 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DT 01-MAY-2000 (TREMREL. 13, Last annotation update)
DE UV-ENDORNUCLEASE.
GN UVSCE.

OS Deinococcus radiodurans.

OC Bacteria; Thermus/deinococcus group; Deinococcales; Deinococcus.

OX NCBI_TaxID=1299;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-KRI;

RA Kitayama S., Kikuchi M., Funayama T., Narumi I., Watanabe H.;

RT "Cloning of structural gene of an alternative incision enzyme for DNA damage in Deinococcus radiodurans.";

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB033747; BAA85759.1; .

KW Endonuclease.

SQ SEQUENCE 305 AA; 33592 MW; B94D33243E2FEA4 CRC64;

Query Match 99.5%; Score 1520; DB 2; Length 305;
Best Local Similarity 99.7%; Pred. No. 3.2e-114;
Matches 293; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QLGLVCLTVGPEVRFTVTLSRYRALSPAEREAKLLDLYSSNIKTLRGAADYCAAHDIRL 60

DB 10 QLGLVCLTVGPEVRFTVTLSRYRALSPAEREAKLLDLYSSNIKTLRGAADYCAAHDIRL 69

QY 61 YRLSSLPFMDLACDGTGAAVLTHLAPQLLEAGHAFTDAGVRLLMHPFEQFVLVNSDRPE 120

DB 70 YRLSSLPFMDLACDGTGAAVLTHLAPQLLEAGHAFTDAGVRLLMHPFEQFVLVNSDRPE 129

QY 121 VRESSVRAMSAHARYMDGLGLARTPNWLLHGGKGGRCALAAALIPDLPVRLRLGLE 180

DB 130 VRESSVRAMSAHARYMDGLGLARTPNWLLHGGKGGRCALAAALIPDLPVRLRLGLE 189

QY 181 NDERAYSPELLPTCEATGTPVFDAAHHVHDKLPDQEDPSVREWLRLARATWQPPFWQ 240

|||||

OS	Bacillus cereus.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;
OT	Bacillus/Staphylococcus group; Bacillus.
OX	NCBI_TaxID=1396;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=ATCC 14579 TYPE STRAIN:
RX	MEDLINE=20055637; PubMed=10589720;
RT	Oksstad O., Gominet M., Purnelle B., Rose M., Lereclus D., Kolsto A.B.
RA	"Sequence analysis of three Bacillus cereus loci under PicR-regulated
RR	genes encoding degradative enzymes and enterotoxin.";
RL	Microbiology 145:3129-3138(1999).
RM	EMBL: AJ243712; CAB69813.1; -
RK	Endonuclease.
RW	SEQUENCE 317 AA: 36990 MW; A6AC35F5800A22C3 CRC64;
SQ	
Query Match 19.5%; Score 297.5; DB 2; Length 317;	
Best Local Similarity 27.1%; Pred. No. 4.2e-16;	
Matches 79; Conservative 64; Mismatches 114; Indels 35; Gaps	
QY	16 RTVLSRYALSPAEREAKLLDLYSSNIKTLRGAADYCAAHDIRLYRSSSLFPML---D 72
DB	: : : : : : : : : : : : : : : : :
DY	21 KTWTFSTFKLKSQREDKLHVIRNOLEHTIRILHYNIAHEIPIYRUSSSIVPLATHPE 80
QY	73 LAGDDTGAAVLTLCAPQLLEAGHAFTDACVRLIMHPHQEIVLNSDRPVESSVRAMSAH 132
DB	: : : : : : : : : : : : : : : : :
DY	81 VFEDYIGV-----FTPWRKIGALIKEHNLRISFHPNQFTLFSDKPHTTTNAITDMIVH 135
QY	133 ARWMDGLGIARTPNLLLLHGKGGRGAEALAA---IPDLPPVRLRGLENDERAYSPA 189
DB	: : : : : : : : : : : : : : : : :
DY	136 YKILDAIGIADSSVIINHVGAYGNKEKAERFHENIKKLPHTHIKKQMILENDKDYTTTA 195
QY	190 ELLPICCATGTPLVFDAHHVVH--DKLPDQEDPSREWVLARATWQ----PPEWQVVH 243
DB	: : : : : : : : : : : : : : : : :
DY	196 ETLSICOKENIPVDFYHHHMAMCLCEPLEELLPAIFE-----TWSHTNISPK---VH 245
QY	244 LSNIGIEGPQRHRSHLI-ADFPSAYADVPO-----IEVAKGKEEAATAAL 287
DB	: : : : : : : : : : : : : : : : :
DY	246 ISSPRSEKEPRAHAEYIDLFTKPFUHIKKHNHFNDIMESKQKDIALFQL 297
RESULT	4
ID	Q9LA4E9 PRELIMINARY; PRT; 325 AA.
OT	Q9LA4E9;
DC	01-OCT-2000 (TEMBLrel. 15, Created)
DT	01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DI	01-MAR-2001 (TEMBLrel. 16, Last annotation update)
DE	POTATIVE UV-ENDONUCLEASE.
DN	YWJD.
GN	Bacillus cereus.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;
OT	Bacillus/Staphylococcus group; Bacillus.
OX	NCBI_TaxID=1396;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=ATCC 14579 TYPE STRAIN:
RX	MEDLINE=20055637; PubMed=10589720;
RT	Oksstad O., Gominet M., Purnelle B., Rose M., Lereclus D., Kolsto A.B.
RA	"Sequence analysis of three Bacillus cereus loci under PicR-regulated
RR	genes encoding degradative enzymes and enterotoxin.";
RL	Microbiology 145:3129-3138(1999).
RM	EMBL: AJ243712; CAB69814.1; -
RK	Endonuclease.
RW	SEQUENCE 325 AA: 37900 MW; 07EFBDDIFA580653 CRC64;
SQ	
Query Match 19.5%; Score 297.5; DB 2; Length 325;	
Best Local Similarity 27.1%; Pred. No. 4.2e-16;	
Matches 79; Conservative 64; Mismatches 114; Indels 35; Gaps	
Y	16 KTVLSRYALSPAEREAKLLDLYSSNIKTLRGAADYCAAHDIRLYRSSSLFPML---D 72


```
RL Nature 406:151-159(2000).
DR EMBL; AE003910; AAF83469.1; -.
KW Complete proteome.
SQ SEQUENCE 437 AA; 48811 MW; ABPBE0825B819DF2 CRC64;

Query Match 6.7%; Score 102.5; DB 2; Length 437;
Best Local Similarity 25.1%; Pred. No. 2.8;
Matches 56; Conservative 30; Mismatches 84; Indels 53; Gaps 12;

QY 43 IKTURGAADYCAAHDIRLYRLSSLPMLDLAGDDTCAAVLTHLAPOLLEAGHA-----F 97
Db 128 LRLALRGSCDGLAA-----MRPLTPFAAGHLWRPLLTTHSRAQLLDYAAQQOHLDWI 176
QY 98 TDAGVRLMLHPEQFIVLNSDRPEVRESSVMSAHARVMDGLGLARTPPWNLILLHGKGG 157
Db 177 EDSNADLRHNRNRLHV-LPLHQWRPORTAVLAR---NAALAANADLL-----224
QY 158 RGAEALALIPDLPVRLRLGLE-NDRAYSAPAE---LLPICEATGTPLVFDAAHHHV- 211
Db 225 -NAEDAVLLPDLLDPDG---ALDINALTAHPARRARLLRAWCARAGAPPLPERGVNII 280
QY 212 HDKLPDQEDPS-----VREHVL-----RARATHQPPWQ 240
Db 281 RELLPARHDSAACTWSHTSIRWRRLRLYLHRPQPPW-PPDWQ 322

RESULT 10
Q9YAQ7 PRELIMINARY; PRT; 313 AA.
ID Q9YAQ7
AC Q9YAQ7;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE HYPOTHETICAL 34.0 KDA PROTEIN APE1886.
GN APE1886.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC Aeropyrum.
OX NCBI_TaxID=56636;
RN SEQUENCE FROM N.A.
RP STRAIN=K1;
RC MEDLINE=99310339; PubMed=10382966;
RA Kwarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jinno K., Takahashi M., Sekine M., Baba S.-I., Anai A., Kosugi H.,
RA Hosoya A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1."
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000062; BAA80891.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 313 AA; 33960 MW; A796B26413607CD0 CRC64;

Query Match 6.5%; Score 98.5; DB 1; Length 313;
Best Local Similarity 25.1%; Pred. No. 3.7;
Matches 65; Conservative 35; Mismatches 84; Indels 75; Gaps 15;

QY 11 PEVRFVTLRSYRALSP-AEREAKLL---DLYSSNKTILRGADYCAAHDIRLYRLSSS 66
Db 55 PELGELVV-----AASPTGEPAARILCVGLHSGGVCRLAHPSE--ASLSLLPYR--EY 105
QY 67 LPFMLDLAGDDTGAAVLTHLAPOLLEAGHAFTDAGVRL---MHP---EOLFVINSDRPE 120
Db 106 EFTLVATGDKPSRIV-HLA-----EAASLLGARGLVVGGPPHHPGYEERLSMLGASRV 160
QY 121 V-RESSVRMSAHARVMDGLGLARTPPWNLILLHGKGGKGR-----158
Db 161 APREAPVLSMS-----IASLVWTPRL--GMREGRFRGEIALLSGSTAWLRRFAG 209
```

```
QY 159 -----GAELAALIPDLPVRLRLGLENDERAYSAPAEELPICEATGTPLVFDAAHHV 211
Db 210 ETEEARGSSSVAYSTPLCLPGALYLSRSRAVHPLMLPLHRTPTPIVF-----262
QY 212 HDKLPDQEDPSVREWLRA 230
Db 263 ---MASVEEPSYRD-VIRS 277

RESULT 11
Q9CG24 PRELIMINARY; PRT; 1486 AA.
ID Q9CG24
AC Q9CG24;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE GLUTAMATE SYNTHASE LARGE SUBUNIT (EC 1.4.1.13).
GN GLTB.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1360;
RN SEQUENCE FROM N.A.
RP STRAIN=IL1403;
RC Bolotin A., Wincker P., Mauger S., Jaillon O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis."
RL Genome Res. 0:0-0(2001).
DR EMBL; AE006360; AAK05384.1; -.
DR InterPro: IPR002489; DUF14.
DR InterPro: IPR002932; Glu_synthase.
DR Pfam: PF01493; DUF14.1.
DR Pfam: PF01645; Glu_synthase; 1.
KW Complete proteome.
SQ SEQUENCE 1486 AA; 164390 MW; 0E9C715CD3C4940D CRC64;

Query Match 6.5%; Score 98.5; DB 2; Length 1486;
Best Local Similarity 23.2%; Pred. No. 29;
Matches 66; Conservative 33; Mismatches 81; Indels 105; Gaps 14;

QY 14 RPTVTLSRYRALSAPAREAKLLDLYSSNKTILRGADYCAAHDIRLY-----61
Db 233 RSTNTFPSSWNRAPPRFLA-----HNGEINTLPGAENMKNVNDIEMYNENSDSAKLE 287
QY 62 -----RLSSSLFPMLDLA-GDDTGAA-----VLTHLAPOLLEAGHAFTD-- 99
Db 288 CMFYFYRNGRELPESLTMTIPEAWGEQGLSPELKAFYEYTAHTAPWGGPAALVFTDCK 347
QY 100 -AGVRL-----LMHPEQFIVLNSD-----RPEVRESSVRMSAHARVMDGLGL 141
Db 348 TVGARLDNRCLRPSRYLVTKDINSILSSSESGVVDIPADEIIEKSV-----LG- 394
QY 142 ARTPNWNLILLHGKGG--RGAEALALIPD-LPDPVRLRLGL-----ENDERAYSAPAE- 190
Db 395 ---PCNMILLVNTDEGKIIRNEEVKSYANKYPYQEFLSAGLKLKLSALTESEKTNPIPSAK 451
QY 191 -----LLPICEATGTPLV---FDAHHVVHDK 214
Db 452 MNTLWLKFGYTDVETVTLVLLPMAESANEPTISMGFDAPLAVLSQ 496

RESULT 12
Q9LV03 PRELIMINARY; PRT; 2216 AA.
ID Q9LV03
AC Q9LV03;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE NADH-DEPENDENT GLUTAMATE SYNTHASE.
```

OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
ON NCBI_TaxID=3702;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=COLUMBIA;
RX MEDLINE=20181125; PubMed=10718197;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC
RT clones.";
RL DNA Res. 7:31-63(2000).
CC -!- COFACTOR: FAD (BY SIMILARITY).
CC -!- SIMILARITY: TO PYRIDINE NUCLEOTIDE-DISULPHIDE OXIDOREDUCTASES
CC CLASS=1.
DR EMBL; AB020754; BAA97323.1; -.
DR InterPro: IPR000759; Adnrdx_reductase.
DR InterPro: IPR002489; DUF14.
DR InterPro: IPR001327; FAD_pyr_redox.
DR InterPro: IPR002932; Glu_pyr_redox.
DR InterPro: IPR000205; NAD_binding.
DR InterPro: IPR00103; Pyridine_redox_2.
DR InterPro: IPR001100; pyr_redox.
DR Pfam; PF01493; DUF14; 1.
DR Pfam; PF01645; Glu_synthase; 1.
DR Pfam; PF00070; pyr_redox; 1.
DR PRINTS; PR00419; ADXRDYASE.
DR PRINTS; PR00368; FADPNR.
DR PRINTS; PR00411; PNDRTASEI.
DR PRINTS; PR00469; PNDRTASEII.
KW FAD; Flavoprotein; Oxidoreductase; Redox-active center.
SQ SEQUENCE 2216 AA; 242877 MW; 2AF258AB8C682371 CRC64;

Query Match 6.3%; Score 96.5; DB 10; Length 2216;
Best Local Similarity 22.0%; Pred. No. 72;
Matches 69; Conservative 54; Mismatches 116; Indels 75; Gaps 17;

QY 14 RFRVTLRYALSPAEREAKLLDLYSSNKTGLRGAADYCAAH--LRLYRLSS-----66
Db 332 RFSNTTPSWDRAP-----NRVLG-HNGEINTLRGNVNRAREGLLKCNELGSKELK 386
QY 67 -LPMLDLAGDGTGAVALTHLAPOLLEAGHAFTDAGVRLMHPEQFIVLNSDR--PEVRE 123
Db 387 KLLPIVDVSSDSGA--FDGVLELLVVRAGRLPEA--VMMNPEAW---QNDKNIDPSRK 439
QY 124 SSVRAMSAHARVMDGLGL-----ARTPWLLLLHGKGGRGGAELAAALIP 167
Db 440 EYEYLSALMEPWDPALISFTDGRYLGATLDNRGLRPGREFYTHSGRVIMASEVG--VV 497
QY 168 DLDPVRLRLGLENDERAYSPAELLPTCEATGTPLVFDAAHHVHVHDKLPDQEDPSPVREW 227
Db 498 DVPEDVMRKGRLN-----PGMMLLVDFEKHIVD-----DDLKQOY 535
QY 228 LRARATQWPQWVHLSNGIEG-PQDRRHSHLIADFPSSAYADVPQ---IEVEAKKEEA 283
Db 536 SLARPYGEWLKQKIEUKDIIESVPEAE-----IA--PSISGVVPLTFLETAKSNDSDS 589
QY 284 IAAL---RLMAPFK 294
Db 590 MESMGHGLLSPLK 603

RESULT 13
Q9X5N3 PRELIMINARY; PRT; 526 AA.
AC Q9X5N3;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE OLIGOPEPTIDE PERMEASE HOMOLOG OPPA.
GN OPPA.
OS Myxococcus xanthus.
OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
OC Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.
ON NCBI_TaxID=34;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=DZFL;
RA Ueki T., Inouye S.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF137082; AAD31006.1; -.
DR InterPro: IPR000914; SBP_bac_5.
DR Pfam; PF00496; SBP_bac_5; 3.
SQ SEQUENCE 525 AA; 58593 MW; AFA585E37B207DB3 CRC64;

Query Match 6.3%; Score 96; DB 2; Length 526;
Best Local Similarity 26.6%; Pred. No. 12;
Matches 58; Conservative 25; Mismatches 81; Indels 54; Gaps 11;

QY 18 VTLSRYALSPAEREAKLLDLYSSNKTGLRGAADY-----52
Db 135 VTLARPSYFLA-RVANVYLFYPAPSADEKGSDEEVDRDRDGRPLALGPYRVERW 193
QY 53 -CAAHDIRLYRLSSSLFPMDLADGDDTGAVALT---HLAPQLF---AGHAFTDAGVRL 104
Db 194 DRAGERVRLVHNPASAFP-RMAEGETVPVITLMKSEIGALYERERVDVFVDSAAAL 252
QY 105 LMH-PEQ-----FIVLNS-----DRPEVRESSVRAHSAHARVMDGLGLARTPNWL 148
Db 253 RTHRPDDLRRREPLLSTYFLAFNTEKAPLDRPEVRNALSALDREA-LLAGLLPAARPSHV 311
QY 149 LLLHGKGGSGGAELAAALIP-LPDPVRLRL-GLENDER 184
Db 312 LLPPELGAATPEQAARLPYAPAEQARAELAGVERPLR 349

RESULT 14
Q9CG88 PRELIMINARY; PRT; 535 AA.
AC Q9CG88;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HYPOTHETICAL PROTEIN YLJF.
GN YLJF.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
ON NCBI_TaxID=1360;
RN [1]
RC SEQUENCE FROM N.A.
RA Holstin A., Wincker P., Mauger S., Jaillon O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis.";
RL Genome Res. 0:0-0(2001).
CC -!- COFACTOR: FAD (BY SIMILARITY).
DR EMBL; AE006350; AAK05277.1; -.
DR InterPro: IPR001327; FAD_pyr_redox.
DR PRINTS; PR00368; FADPNR.
KW Complete proteome; FAD; Flavoprotein; Hypothetical protein;
KW Oxidoreductase.
SQ SEQUENCE 535 AA; 58670 MW; 116102ABB9FAEC95 CRC64;

Query Match 6.3%; Score 96; DB 2; Length 535;
Best Local Similarity 24.4%; Pred. No. 12;
Matches 68; Conservative 36; Mismatches 87; Indels 88; Gaps 17;

QY 11 PEVFRVTVLSRYALSPAEREAKLLD--LYSSNKTGLRGAADYCAAHDIRL---YRLS 64

```
Db 14 PVSKVSLVKKLRL-----QESDLDYRIYKESIDARHGEIDFYTVDIKLDKDEARIL 68
Qy 65 S-----SLFPMLDLAGDDTGAAVLTHLAPQLLEAGHAFTDAGVRLMLHPEQFIVLNSDR 118
Db 69 SKKIKNVSLAPELDYVNPVIGSKKMLH-RPVTI---GFGPAG---MFAALLLAQNGYR 119
Qy 119 PEVRESSVRAMSAHARVMDGLGLARTPWNLILLH-----GKGGRG-----AELAA 164
Db 120 PIVLERG-QAVDERVKSIDKF-----W-----LEGKLNPKSNVOFGEGGAGTFSDGKLT 168
Qy 165 LIPDLPDVRRLGLRGLENDERAYSPAELLPICEATGTPLVEDAHHHVVDHKKLPD----- 217
Db 169 RVRDL----RGRKVLEEFVQAGAPEDIL-----YKAHPHVGTDLRLDIVKNIRK 213
Qy 218 -----QEDPSVREWVLRARATWQPPVQVHLSNG 247
Db 214 QIIELGGEVHFDAQVEEFLI-----EDEELQAVKLADG 246
```

RESULT 15

```
O80962
ID O80962 PRELIMINARY; PRT; 791 AA.
AC O80962;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE T16B24.17 PROTEIN.
GN T16B24.17.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Rounsley S.D., Kaul S., Lin X., Ketchum K.A., Crosby M.L.,
RA Brandon R.C., Sykes S.M., Mason T.M., Kerlavage A.R., Adams M.D.,
RA Somerville C.R., Venter J.C.;
RT "Arabidopsis thaliana chromosome II BAC T16B24 genomic sequence.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC004697; AAC28988.1;
DR InterPro; IPR000209; Peptidase_S8.
DR PROSITE; PS00136; SUBTILASE_ASP; UNKNOWN_1.
SQ SEQUENCE 791 AA; 88694 MW; C771BC05757D2863 CRC64;
```

```
Query Match 6.3%; Score 96; DB 10; Length 791;
Best Local Similarity 24.1%; Pred. No. 20;
Matches 83; Conservative 30; Mismatches 107; Indels 124; Gaps 18;

Qy 36 LDLYSSNITLIRGAADYCAAHDIRLYRLSSLLFPMLDLAGDDTGAAVLTHLAP-----QL 90
Db 114 VDLFMADVKALETYAGY-----FYSLSKMSRPLPEVYDSQAVDYFNCRPHVWAFRL 166
Qy 91 LEAGHAFTDAGVRL-----LMHPEQ-----FIVLN---SDRPEV 121
Db 167 LEVFAFTIAAIRLRTSASVKGSLEASQIFGMVKETMLHLGPTFTKVGQSLSRTPDI 226
Qy 122 --RESSVRAMSAHARVMDGLGLARTPW--NLLLLHGGKGG----- 157
Db 227 IGTEISKALSELHDI-----PPFPWPEAVKILEGELGGVSEFFSOFQSVTAASFG 280
Qy 158 -----RGAELAALI--POLPDVPR-----LRGL-----ENDERAYS----- 187
Db 281 QYVGRGTLDGADVAVKVGQRPDLRHAVLRDIVILRLGLVLRKAKRENDIRVYADELCMG 340
Qy 188 -PAELLPICEATGTPLVDFAHHVVDHKLDPQEDPSVREWVLRARATWQPPVQVHLSN 246
Db 341 LAGELDTLEANASEFQEAHAREFSYIRPKVYQHLTRKRVLTM-----EMVVG----- 389
Qy 247 GIEGQDRRHSHLIADFTFSAYADVQIEVEAKGKEEAIAALRLM 290
```

```
Db 390 --ESPTD-----LLA-ISSGYSD---HDNESHEKQKIEARRRELL 422

Search completed: January 15, 2002, 14:06:33
Job time: 972 sec
```


No PA

How does this work?
how to handle?

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:52:37 ; Search time 74.94 Seconds
(without alignments)
298.843 Million cell updates/sec

Title: US-09-724-296-39
Perfect score: 1527
Sequence: 1 QLGLVCLTVGPEVRFRTVTL.....VEAKGKEEIAALRLMAPFK 294

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1520	99.5	326	2 C75350	probable UV damage
2	303	19.8	320	2 S55418	UV-endonuclease ho
3	255.5	16.7	656	2 S55262	UV-endonuclease -
4	250	16.4	322	2 E83974	hypothetical prote
5	230.5	15.1	599	2 S71134	UV-endonuclease -
6	102.5	6.7	437	2 B82778	cell cycle protein
7	98.5	6.5	313	2 F72375	hypothetical prote
8	98.5	6.5	1486	2 F86785	glutamate synthase
9	96	6.3	535	2 C82772	conserved hypotet
10	96	6.3	791	2 T02583	probable ABC trans
11	95	6.2	368	2 T36004	hypothetical prote
12	95	6.2	515	2 H75579	hypothetical prote
13	94.5	6.2	1809	2 T17403	pyochelin syntheta
14	94.5	6.2	1809	2 C83118	pyochelin syntheta
15	94	6.2	898	2 T37037	pyruvate phosphata
16	93	6.1	199	2 JU0052	puromycin N-acetyl
17	93	6.1	603	2 A75373	probable N-acetyl
18	92	6.0	332	2 T17427	probable FK506 oxl
19	92	6.0	931	2 G82965	conserved hypotet
20	91	6.0	344	2 E83562	N-acetyl-gamma-glu
21	90.5	5.9	457	2 T10360	late expression fa
22	90.5	5.9	842	2 C83177	probable phosphotr
23	90	5.9	488	2 S27652	probable aldehyde
24	90	5.9	608	1 S01040	glutamine--fructos
25	90	5.9	3413	2 T17467	rifamycin polyketi
26	89.5	5.9	319	2 H83253	conserved hypotet
27	89.5	5.9	390	2 D83853	cystathionine beta
28	89.5	5.9	397	2 T35289	probable multifunc
29	89.5	5.9	1602	2 H70984	probable polyketid

30 89 5.8 447 2 H72516
31 88.5 5.8 650 2 T17482
32 88.5 5.8 1221 2 T23472
33 88 5.8 253 2 A75252
34 88 5.8 352 2 G71511
35 88 5.8 537 2 I49135
36 88 5.8 707 2 G96520
37 87.5 5.7 439 2 F75306
38 87 5.7 558 2 E84324
39 87 5.7 10223 2 T30225
40 86.5 5.7 415 2 T33536
41 86 5.6 636 2 H83040
42 86 5.6 1110 1 A70652
43 85.5 5.6 653 2 A46362
44 85 5.6 360 2 B75319
45 84.5 5.5 417 2 F81105

ALIGNMENTS

RESULT 1
C75350
probable UV damage endonuclease - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: C75350
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioreistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: C75350
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-326 <WHI>
A:Cross-references: GB:AE000222; GB:AE000513; NID:g6459590; PIDN:AAF11370.1; PID:g645
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1819
A:Map position: 1

Query Match 99.5% Score 1520; DB 2; Length 326;
Best Local Similarity 99.7% Pred. No. 2.3e-122;
Matches 293; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 QLGLVCLTVGPEVRFRTVTLRYRALSPAEREAKLLDLYSSNIKTILRGADYCAAHDIRL 60
Db 31 QLGLVCLTVGPEVRFRTVTLRYRALSPAEREAKLLDLYSSNIKTILRGADYCAAHDIRL 90
QY 61 YRLSSSLFPMLDLAGDDTCAAVLTHLAPOLLGAGHAFTDAGVRLLMHPQFIVLNSDRPE 120
Db 91 YRLSSSLFPMLDLAGDDTCAAVLTHLAPOLLGAGHAFTDAGVRLLMHPQFIVLNSDRPE 150
QY 121 VRESSVRAMSAHARVMDGLGLARTPNWLLHGGKGGRAELAAALIPDLDPVRLRLGLE 180
Db 151 VRESSVRAMSAHARVMDGLGLARTPNWLLHGGKGGRAELAAALIPDLDPVRLRLGLE 210
QY 181 NDERAYSPAELLPTICATGTPLVFDAAHHVHVHDKLPDQEDPSVREWLRLARATWOPPEWQ 240
Db 211 NDERAYSPAELLPTICATGTPLVFDAAHHVHVHDKLPDQEDPSVREWLRLARATWOPPEWQ 270
QY 241 VVHLSNGTFPGDQRHSHLIADFPSPAYADVPOTEVEAKGKEEIAALRLMAPFK 294
Db 271 VVHLSNGTFPGDQRHSHLIADFPSPAYADVPOTEVEAKGKEEIAALRLMAPFK 324

RESULT 2
S55418
UV-endonuclease homolog ywjd - Bacillus subtilis
C:Species: Bacillus subtilis

Query Match 6.5%; Score 98.5; DB 2; Length 313;
Best Local Similarity 25.1%, Pred. No. 0.64;
Matches 65; Conservative 35; Mismatches 84; Indels 75; Gaps 15;

QY 11 PEVRFRTVTLSRYRALSP-AERAKLLD---DLYSNNIKTLRGAADYCAAHDIRLYLSSS 66
||| : | | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 55 PELKGELV-----AASPTGEPAARTLCVGVLHSGGVCRLAHPS--ASLSILPYR--EY 105

QY 67 LFPMLDLAGDDTGAAVLTHLPOLLEAGHAFTDAGVRLL---MHP---EQFIVLNSDRPE 120
||| : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 106 EFTLVAFDTGDKPDSRIV-HLA---EAASLLGARGLYVGPMPHGGYBERLSMEGASRVE 160
||| : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |

QY 121 V-RESSVRAMSAHARYMDGLIGLARTPNWLLLHGKGGKGR----- 158
||| : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |

Db 161 APREAPVLSMS-----IASLVWTPRLL-GMREGFRCEIEALSGSTAWLRERFAG 209

QY 159 -----GAELAALIPLDPVRLRLGLENDERAYSPAELLPTCEATGTPLVFDAHHHV 211
||| : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |

Db 210 EIEBARGSSVEAYSTPLGPLGALYLXRSQRSRAVHPLEMLPLHRTPTPIVF----- 262

QY 212 HDKLDPQEDPSVREWVLA 230

Db 263 ---MASVEPSYRD-VIRS 277

RESULT 8
F86785
glutamate synthase (NADPH) (EC 1.4.1.13) large chain [imported] - Lactococcus lactis sub
C:Species: Lactococcus lactis subsp. lactis
C>Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 31-Mar-2001
C:Accession: F86785
R:Botolin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrlich
Genome Res. in press, 2001
A:Title: The complete genome sequence of the lactic acid bacterium.
A:Reference number: A86625
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1486 <STO>
A:Cross-references: GB:AE005176; NID:g12724262; PIDN:AAK05384.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: gltB
C:Superfamily: glutamate synthase (NADPH)
C:Keywords: oxidoreductase

Query Match 6.5%; Score 98.5; DB 2; Length 1486;
Best Local Similarity 23.2%, Pred. No. 5;
Matches 66; Conservative 33; Mismatches 81; Indels 105; Gaps 14;

QY 14 RFTVTLSRYRALSPAEREAKLLLDYSSNIKTLRGAADYCAAHDIRLY----- 61
||| : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 233 RFSTNTFPSWNRAQPFRFLA-----HNGEINTLRGAENMKVMVDIEMYNENSDSA KLEN 287

QY 62 -----RLSSSLFPMPLDLA-GDDTGAA-----VLTHLPQLLEAGHAFTD-- 99

Db 288 CMFEFYRNCRPELSLLTWIPFAWEQGTGLSPKLFAYESTAHIAPDWDGPAALVFDGK 347

QY 100 -AGVRL-----LMHQEFIVLNSD-----RPVESSVRAMSAHARYMDGLGL 141
||| : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |

Db 348 TVCARLDNRGLRPSRYLTVTKNSIISSSEGVDPIDADEITKSV-----LG- 394

QY 142 ARTPNWLLLHGKGG--RGAEALALPD-LPDVRLRLGL-----ENDERAYSPA E- 190
||| : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |

Db 395 ---PGNMILLVNDEGKIIRNEEVKSYYANKPYOEFLSAGLKLSALTESKTNPIPSAK 451

QY 191 -----LLPTCEATGTPLV---FDAHHHVHDK 214
||| : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |

Db 452 MNTLWKFGYTDEVIRTVLLPMAESANEPTISMGFDAPLAVLSDQ 496

Query Match 6.3%; Score 96; DB 2; Length 535;
Best Local Similarity 24.4%, Pred. No. 2.1;
Matches 68; Conservative 36; Mismatches 87; Indels 88; Gaps 17;

QY 11 PEVRFRTVTLSRYRALSPAEREAKLLD---LYSSNIKTLRGAADYCAAHDIRL---YRLS 64
||| : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 14 PVSKVKSLVLKKLRL----QESDLLDYRYKESIDARHGERIDFIYTDIKLDEARIL 68

QY 65 S-----SLFPMDLDLAGDDTGAAVLTHLPOLLEAGHAFTDAGVRLLMHPEQFIVLNSDR 118
||| : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |

Db 69 SKKIKNVSLAPELDYVNPVIGSKKMLH-RPVVI---GFPGAG---MFALLLAQNGYR 119

QY 119 PVESSVRAMSAHARYMDGLGLARTPNWLLLHG-----GKGGRG-----AELAA 164
||| : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |

Db 120 PIVLERG-QAVDERVKSIDKF-----W-----LEGKLNPKSNVQFEGGAGTSDGKLT S 168

QY 165 LIPDLDPDVRLRLGLENDERAYSPAELLPTCEATGTPLVFDAHHHVHDKLPD----- 217
||| : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |

Db 169 RVRDLL---RGRKVLKEEFVQACAPEDIL-----YKAHPHVGTDLLRIDVKNIRK 213

QY 218 -----QQDPPSVREWVLRARATWPPEVQVVHLNSG 247

Db 214 QIIELGGEVHFDAQVEEFLI-----EDELOAVKVLADG 246

RESULT 10
T02583
probable ABC transporter At2g39190 [imported] - Arabidopsis thaliana
N:Alternate names: hypothetical protein T16P24.17
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C:Accession: T02583; C84814
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Syke
submitted to the EMBL Data Library, August 1998
A:Description: Arabidopsis thaliana chromosome II BAC T16B24 genomic sequence.
A:Reference number: Z14679
A:Accession: T02583
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-791 <ROU>
A:Cross-references: EMBL:AC004697; NID:g3402671; PID:g3402685
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffatt, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon,
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: C84814
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-791 <STO>
A:Cross-references: GB:AE002093; NID:g3402685; PIDN:AAC28988.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g39190; T16B24.17

A:Map position: 2

A:Introns: 110/1; 168/3; 216/1; 233/3; 281/3; 316/3; 358/3; 424/3; 460/2; 514/3; 536/3;

Query Match 6.3%; Score 96; DB 2; Length 791;

Best Local Similarity 24.1%; Pred. No. 3.6;

Matches 83; Conservative 30; Mismatches 107; Indels 124; Gaps 18;

QY 36 LDLYSSNKTILRGADYCAAHDIRLYRLSSSLFPMFLDLAGDDTGAANVLTHLAP-----QL 90

DB 114 VDLFMADYKALETVAGY-----FYSLSKMSRPLPEVYDSQAVADYFNCRPHWAFRL 166

QY 91 LEAGHAFTDAGVRL-----LMHPEQ-----FIVLN-----SDPEV 121

DB 167 LEVFAFTIAIRLTSASVSGKSEASEQIFGMVLKETMLHLGPTFIKVGQSLSRPTDI 226

QY 122 --RESSVRMSAHARVMDGLGLARTPW--NLLLHGKGGK-----FIVLN-----SDPEV 157

DB 227 IGTEISKALSELHDRI-----PPWPPEAVKILLEGELGGPVESFFSQFVAAASFG 280

QY 158 -----RGAEALALI--PDLDPDVR-----LRIGL-----ENDERAYS----- 187

DB 281 QYVGRITLDGADVAVKVGQRPDLRHAVLRDIYILRLGLGVLRKVKAKRENDIRVYADELGMG 340

QY 188 -PAELLPICEATGTPLVFDAHHVVDKLPDOEDPSVREWVLRARATWQPEWQVHLSN 246

DB 341 LAGELDFTLEANASEFOEHAHFARSYIRVPKYQHLTRKRVLTM-----EWMVG----- 389

QY 247 GIEGQDRRHSHLIADFPFSAYADVPOIEVEAKGEEAIAALRLM 290

DB 390 --ESPTD-----LLA-ISSGSD---HDNESHEKQIEARRLL 422

RESULT 11

T36004

hypothetical protein SCC22.14c - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 08-Sep-2000

C:Accession: T36004

R:Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, July 1999

A:Reference number: 221574

A:Accession: T36004

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-368 <SEE>

A:Cross-references: EMBL:AL096839; PIDN:CAB50757.1; GSPDB:GN00070; SCOREDB:SCC22.14c

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOREDB:SCC22.14c

C:Superfamily: Streptomyces coelicolor hypothetical protein SCC22.14c

Query Match 6.2%; Score 95; DB 2; Length 368;

Best Local Similarity 26.0%; Pred. No. 1.6;

Matches 81; Conservative 27; Mismatches 103; Indels 100; Gaps 18;

QY 11 PEVFRFTVTLRSYRALSPAEREAKLLDLYSSNIKT-----LRGAADYCAAH----- 57

DB 66 PPARV-----LARRRELGVMEADRRL--LRGSGITTYLVLDAGLPGDLTGPTMETATAADAT 119

QY 58 ---IRLYRLSSSLFPMFLDLAGDDTGAANVLTHLAPOLLGAGH-----AFTD-AGVR--L 104

DB 120 REIVRLELLAEQV-----ADTSGTVESFLA-NLAEAVHGAANAVAFSTVAGVRHGL 170

QY 105 LMHPQFIVLNSDRPEVRESSVRMSAHARVMDGLGLARTPNWLLLLHGGKGGCAELAA 164

DB 171 ALAPEP-----PGQEVGGAARWUT-----GREVGG 197

QY 165 LIPDLDPVRLRLGLENDERAYSAPLELLPICEATGTPLVFDAHHVVDKLP-DOEDPSV 223

DB 198 ---ELSDPVLLR-----HLLWIAVASGLPL--QLHAGLGEPLGRIDRTDPVL 239

QY 224 RENWLRARATWQPEWQVHLSNGIEGPODRRHSHLIADFPFSAYADVPOIEVEAKGKEA 283

DB 240 LTFVTRTAGLGTGTD----LVLLHG--PYHRHAHLAGVFPHYVADSGAALVVRTGARAAT 293

QY 284 IAALRL-MAPF 293

DB 294 VLAFILELAPF 304

RESULT 12

H75579

hypothetical protein - Deinococcus radiodurans (strain RL)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000

C:Accession: H75579

R:White, O.; Eisen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.

, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans RL.

A:Reference number: A75250; MUID:20036896

A:Accession: H75579

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-515 <WHL>

A:Cross-references: GB:AE001863; GB:AE001825; NID:g6460670; PIDN:AAF12492.1; PID:g646

A:Experimental source: strain RL

C:Genetics:

A:Gene: DRA0265

A:Map position: 2

Query Match 6.2%; Score 95; DB 2; Length 515;

Best Local Similarity 24.5%; Pred. No. 2.5;

Matches 48; Conservative 16; Mismatches 86; Indels 46; Gaps 7;

QY 78 TGAAYLTHLAPQLLEAGHAFTDAGVRLLMHPEQFIVLNSDRPEVRESSVRMSAHARVMD 137

DB 109 TAFALARMVAELPE-----LQRRHFE-----GGDLPEALSGVOTLMTVPLRAQD 154

QY 138 GLGLARTPNWLLLL-----HGGKGGCAELAAIIPDLDPVRLRLGLENDERAYS 189

DB 155 LLELAGEAWQLWQLARQRPRAHSAQAQRARERVLGGQVDPDRVDMNRTLDLWGRGEFP- 213

QY 190 ELLPICEATGTPLVFDAHHVVDKLPDQEDPSVREWVLRARATWQPEWQVHLSNGIE 249

DB 214 -----HHVALD-LPAPQPPA---LAALRELWQAMEEAAAQLAPGPE 251

QY 250 GPQ-DRHSHLIADFP 264

DB 252 RDEVORRFARALAEFP 267

RESULT 13

T17403

pyochelin synthetase (EC 6.3.2.-) - Pseudomonas aeruginosa

C:Species: Pseudomonas aeruginosa

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000

C:Accession: T17403

R:Reimann, C.; Serino, L.; Beyeler, M.; Haas, D.

Microbiology 144, 3135-3148, 1998

A:Title: Dihydroaeruginic acid synthetase and pyochelin synthetase, products of the

A:Reference number: Z18770; MUID:99061207

A:Accession: T17403

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1809 <REI>

A:Cross-references: EMBL:AF074705; NID:g3386352; PID:g3386354; PIDN:AAC83657.1

C:Genetics:

A:Gene: pchF

C:Function:

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 14:06:28 ; Search time 134.53 Seconds
(without alignments)
630.625 Million cell updates/sec

Title: US-09-724-296-38
Perfect score: 3055
Sequence: 1 MTTTGLSLSLDRGAAPTV.....HKLKHYHDWLNWENHEKLSLS 580

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3055	100.0	580	4 095243	095243 homo sapien
2	1821.5	59.6	554	11 0922D7	Q922d7 mus musculu
3	880.5	28.8	416	13 0919F1	Q919f1 gallus gall
4	344	11.3	419	10 09SFC1	Q9Sfc1 arabidopsis
5	220.5	7.2	467	13 09IGC6	Q9Ygc6 xenopus lae
6	202.5	6.6	344	13 042403	042403 gallus gall
7	146	4.8	702	10 09SFC2	Q9Sfc2 arabidopsis
8	143.5	4.7	782	5 025875	Q25875 plasmodium
9	141.5	4.6	782	5 025730	Q25730 plasmodium
10	141.5	4.6	782	5 026007	Q26007 plasmodium
11	140.5	4.6	782	5 09U430	Q9u430 plasmodium
12	140.5	4.6	782	5 09U414	Q9u414 plasmodium
13	138.5	4.5	782	5 09U429	Q9u429 plasmodium
14	137.5	4.5	782	5 026104	Q26104 plasmodium
15	137.5	4.5	2209	5 09U0G6	Q9u0g6 plasmodium
16	136.5	4.5	782	5 09U431	Q9u431 plasmodium
17	136.5	4.5	1359	5 002061	002061 caenorhabdi
18	135.5	4.4	282	13 09PUM9	Q9pum9 xenopus lae
19	135.5	4.4	755	4 043719	043719 homo sapien

20	135.5	4.4	1058	5 09V433	09v433 drosophila
21	135	4.4	822	3 09USH9	09ush9 schizosacch
22	133.5	4.4	1359	5 09U7E0	09u7e0 caenorhabdi
23	132.5	4.3	286	1 09YDP0	09ydp0 aeropyrum p
24	132	4.3	1819	2 09ZLV0	09zlv0 helicobacte
25	131.5	4.3	230	1 09P9L6	09p9l6 pyrobaculum
26	131.5	4.3	285	11 09Z2D8	09z2d8 mus musculu
27	131.5	4.3	291	4 095983	095983 homo sapien
28	131	4.3	754	4 099730	099730 homo sapien
29	130.5	4.3	411	4 09UBH5	09ubh5 homo sapien
30	130	4.3	537	5 09VZT4	09vzt4 drosophila
31	130	4.3	1927	2 025262	025262 helicobacte
32	129.5	4.2	3122	11 09JMD6	09jmd6 mus musculu
33	129	4.2	1163	4 09UHB7	09uhb7 homo sapien
34	128.5	4.2	946	10 09C9D8	09c9d8 arabidopsis
35	128	4.2	853	10 09LFE4	09lfe4 arabidopsis
36	127.5	4.2	1058	5 09U5E0	09u5e0 drosophila
37	127	4.2	648	10 09LQV6	09lqv6 arabidopsis
38	127	4.2	970	10 09AYF2	09ayf2 oryza sativ
39	127	4.2	1723	2 09JMX8	09jmx8 helicobacte
40	126.5	4.1	1183	5 09W3D1	09w3d1 drosophila
41	126	4.1	303	13 09PUM8	09pum8 xenopus lae
42	126	4.1	561	10 09C6S8	09c6s8 arabidopsis
43	126	4.1	881	5 09BKN8	09bkn8 plasmodium
44	125.5	4.1	302	4 09UIS8	09uis8 homo sapien
45	125	4.1	873	11 09WVC9	09wvc9 mus musculu

ALIGNMENTS

RESULT	1				
095243					
ID	095243	PRELIMINARY;	PRT:	580	AA.
AC	095243;				
DT	01-MAY-1999	(TREMBLrel. 10, Created)			
DT	01-MAY-1999	(TREMBLrel. 10, Last sequence update)			
DE	01-JUN-2001	(TREMBLrel. 17, Last annotation update)			
DE	METHYL-CPG BINDING PROTEIN MED4.				
GN	MBD4 OR MED1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=98449942; PubMed=9774669;				
RA	Hendrich B., Bird A.;				
RT	"Identification and characterization of a family of mammalian methyl-				
RT	CpG binding proteins";				
RL	Mol. Cell. Biol. 18:6538-6547(1998).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=FETAL BRAIN;				
RX	MEDLINE=99199294; PubMed=10097147;				
RA	Bellacosa A., Cicchillitti L., Schepis F., Riccio A., Yeung A.T.,				
RA	Matsumoto Y., Golemis E.A., Genuardi M., Neri G.;				
RT	"MED1, a novel human methyl-CpG-binding endonuclease, interacts with				
RT	DNA mismatch repair protein MLH1";				
RL	Proc. Natl. Acad. Sci. U.S.A. 96:3969-3974(1999).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=99373255; PubMed=1041743;				
RA	Hendrich B., Abbott C., McQueen H., Chambers D., Cross S., Bird A.;				
RT	"Genomic structure and chromosomal mapping of the murine and human				
RT	mbd1, mbd2, mbd3, and mbd4 genes.";				
RL	Mamm. Genome 10:906-912(1999).				
DR	EMBL; AF072250; AAC68879.1; -				
DR	EMBL; AF114784; AAD22195.1; -				
DR	EMBL; AF120999; AAD50374.1; -				
DR	EMBL; AF120997; AAD50374.1; JOINED.				
DR	EMBL; AF120998; AAD50374.1; JOINED.				
DR	InterPro; IPR001739; MBD.				

DR InterPro: IPR003265; Endo_3c.
 DR Pfam: PF00730; HHH-GPD; 1.
 DR Pfam: PF01429; MBD; 1.
 DR SMART: SM00391; MBD; 1.
 KW Endonuclease
 SQ SEQUENCE 580 AA; 66050 MW; BF16FB21A34B8E5F CRC64;

Query Match 100.0%; Score 3055; DB 4; Length 580;
 Best Local Similarity 100.0%; Pred. No. 1.9e-219;
 Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGTGLESLSLGRGAATVTSERLVPDPNDLRKEDVAMELERVGEDEQMMIKRSSE 60
 DB 1 MGTGLESLSLGRGAATVTSERLVPDPNDLRKEDVAMELERVGEDEQMMIKRSSE 60

QY 61 CNPLQLEPIASQAFCATAGTECRKSVPCGWERYVVKORLFGKTAGTDFVYFISPOGLKFRS 120
 DB 61 CNPLQLEPIASQAFCATAGTECRKSVPCGWERYVVKORLFGKTAGTDFVYFISPOGLKFRS 120

QY 121 KSSLANYLHKNGETSLKPEDFTVLSKRGTKSRKDKCSMAALTSHLQNSNNWNLT 180
 DB 121 KSSLANYLHKNGETSLKPEDFTVLSKRGTKSRKDKCSMAALTSHLQNSNNWNLT 180

QY 181 RSKCKKDVMPSPSSSELSQESRGLSNFTSTHLLKDEGVDVNFVRKPKGKVTILKG 240
 DB 181 RSKCKKDVMPSPSSSELSQESRGLSNFTSTHLLKDEGVDVNFVRKPKGKVTILKG 240

QY 241 IPKTKKGCRCSCGFGVQSDSKRESVNCADAESEPVAAQSOLDRTVCIISDAGACGETL 300
 DB 241 IPKTKKGCRCSCGFGVQSDSKRESVNCADAESEPVAAQSOLDRTVCIISDAGACGETL 300

QY 301 SVTSEENSLVKKERSLSSGNSFCSEQKTSGLINKFCSEAKDSEHNEKYEDTFLESEETGT 360
 DB 301 SVTSEENSLVKKERSLSSGNSFCSEQKTSGLINKFCSEAKDSEHNEKYEDTFLESEETGT 360

QY 361 KVEVVERKEHLHTDILKRGSEMDNCSPTKDFTEGKIFQEDTIPRTQIERKTSFYSS 420
 DB 361 KVEVVERKEHLHTDILKRGSEMDNCSPTKDFTEGKIFQEDTIPRTQIERKTSFYSS 420

QY 421 KYNKEALSPRRKAPKKTWPPSPFNVLQETLFDHPWKLLIATIFLNRTSGKMAIPVLWK 480
 DB 421 KYNKEALSPRRKAPKKTWPPSPFNVLQETLFDHPWKLLIATIFLNRTSGKMAIPVLWK 480

QY 481 FLEKYPSEAVARTADWRDVSSELLKPLGLYDLRAKTIKFSDEYLTWKQYPIELHGIGKY 540
 DB 481 FLEKYPSEAVARTADWRDVSSELLKPLGLYDLRAKTIKFSDEYLTWKQYPIELHGIGKY 540

QY 541 GNDYSRIFCVNWKQVHPDHKLKHYHDLWLNWENHEKLSLS 580
 DB 541 GNDYSRIFCVNWKQVHPDHKLKHYHDLWLNWENHEKLSLS 580

RESULT 2
 Q92D7
 ID Q92D7 PRELIMINARY; PRT; 554 AA.
 AC Q92D7
 DT 01-MAY-1999 (TremBLrel. 10, Created)
 DT 01-MAY-1999 (TremBLrel. 10, Last sequence update)
 DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
 DE METHYL-CPG BINDING PROTEIN MBD4.
 GN MBD4.
 OS Mus musculus (Mouse), and
 OS Mus musculus domesticus (western European house mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090; 10092;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98449942; PubMed=9774669;
 RA Hendrich B., Bird A.;
 RT "Identification and characterization of a family of mammalian methyl-
 CpG binding proteins.";

RL Mol. Cell. Biol. 18:6538-6547 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129;
 RX MEDLINE=99373255; PubMed=10441743;
 RA Hendrich B., Abbott C., McQueen H., Chambers D., Cross S., Bird A.;
 RT "Genomic structure and chromosomal mapping of the murine and human
 Mbd1, mbd2, mbd3, and mbd4 genes.";
 RL Mamm. Genome 10:906-912 (1999).
 DR EMBL: AF072249; AAC68878.1; -.
 DR EMBL: AF120996; AAD56595.1; -.
 DR MGD: MGI:1333850; Mbd4.
 DR InterPro: IPR003265; Endo_3c.
 DR InterPro: IPR001739; MBD.
 DR InterPro: IPR002052; N6.Mtase.
 DR Pfam: PF00730; HHH-GPD; 1.
 DR Pfam: PF01429; MBD; 1.
 DR SMART: SM00391; MBD; 1.
 DR PROSITE: PS00092; N6.MTASE; UNKNOWNL.
 SQ SEQUENCE 554 AA; 62577 MW; 792D37CB180291F5 CRC64;

Query Match 59.6%; Score 1821.5; DB 11; Length 554;
 Best Local Similarity 66.2%; Pred. No. 1.4e-127;
 Matches 384; Conservative 49; Mismatches 116; Indels 31; Gaps 11;

QY 6 LESLSLGD---RGAATVTSERLVPDPNDLRKEDVAMELERVGEDEQMMIKRSSECN 62
 DB 1 MESPNLGDNRVGC-----ESLVPDPWDRCKEDIAVGLGVGVEDGKDLVI--SSERS 50

QY 63 PLQLEPIASQAFCATAGTECRKSVPCGWERYVVKORLFGKTAGTDFVYFISPOGLKFRSKS 122
 DB 51 SLQLEPIAST--LSSTTATEGHKVPVPCGWERYVVKORLFGKTAGTDFVYFISPOGLKFRSKR 109

QY 123 SLANYLHKNGETSLKPEDFTVLSKRGTKSRKDKCSMAALTSHLQNSNNWNLT 182
 DB 110 SLANYLHKNGETSLKPEDFTVLSKRGTKSRKDKCSMAALTSHLQNSNNWNLT 169

QY 183 KCKKDVMPSPSSSELSQESRGLSNFTSTHLLKDEGVDVNFVRKPKGKVTILKG 242
 DB 170 KCKKDVMPSPSSSELSQESRGLSNFTSTHLLKDEGVDVNFVRKPKGKVTILKG 229

QY 243 IKTKKGCRCSCGFGVQSDSKRESVNCADAESEPVAAQSOLDRTVCIISDAGACGETLSV 302
 DB 230 SQTOKCRKSLLESQNRKRASVVKVQKADRELVPQESQLNRTLCPADACA--RETGL 288

QY 303 TSEENSLVKKERSLSSGNSFCSEQKTSGLINKFCSEAKDSEHNEKYEDTFLESEETGT 362
 DB 289 AGE-----EKSPGLDLCFIQVTSGLTNKFSHTEAAGEANR--EQTFLESEETRSK- 338

QY 363 EVVERK--EHLHTDILKRGSEMDNCSPTKDFTEGKIFQEDTIPRTQIERKTSFYSS 420
 DB 339 --GDRKGEAHLHTGVLDGSEMP--SCSQAKKHFTSE--TFQEDSIPRTQVEKRKTSFYSS 394

QY 421 KYNKEALSPRRKAPKKTWPPSPFNVLQETLFDHPWKLLIATIFLNRTSGKMAIPVLWK 480
 DB 395 KYNKEALSPRRKAPKKTWPPSPFNVLQETLFDHPWKLLIATIFLNRTSGKMAIPVLWE 454

QY 481 FLEKYPSEAVARTADWRDVSSELLKPLGLYDLRAKTIKFSDEYLTWKQYPIELHGIGKY 540
 DB 455 FLEKYPSEAVARTADWRDVSSELLKPLGLYDLRAKTIKFSDEYLTWKQYPIELHGIGKY 514

QY 541 GNDYSRIFCVNWKQVHPDHKLKHYHDLWLNWENHEKLSLS 580
 DB 515 GNDYSRIFCVNWKQVHPDHKLKHYHDLWLNWENHEKLSLS 554

RESULT 3
 Q919FI
 ID Q919FI PRELIMINARY; PRT; 416 AA.
 AC Q919FI;
 DT 01-OCT-2000 (TremBLrel. 15, Created)
 DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)


```

Db 421 EVISRDGHMKLQTSVFLASIDQRSERAGSACTAL-----VAVIADWFQKNGN 469
QY 503 LKPLGLYDLRAKTIIVKFSDEYLTQWKYPYIELHIGIKYGNDSYRIFCVNE-WKOVHPEDH 561
Db 470 LMP-----IKSOFDSLIRE-----GSLEWRNLNENETIMOKFPDKH 505

RESULT 8
Q25875
ID Q25875 PRELIMINARY; PRT: 782 AA.
AC Q25875;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE P86 RHOPTRY PRECURSOR PROTEIN.
GN P82.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HONDURAS I CDC;
RX MEDLINE=90348711; PubMed=2200961;
RA Ridley R.G., Takacs B., Lahm H.W., Delves C.J., Goman M., Certa V.,
RA Matile H., Woollett G.R., Scaife J.G.;
RT "Characterisation and sequence of a protective rhoptry antigen from
RT Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 41:125-134 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HONDURAS I CDC;
RX MEDLINE=92244303; PubMed=1574089;
RA Howard R.F.;
RT "The sequence of the p82 rhoptry protein is highly conserved between
RT two Plasmodium falciparum isolates.";
RL Mol. Biochem. Parasitol. 51:327-330 (1992).
DR EMBL; M80807; AAA29717.1;
SQ SEQUENCE 782 AA; 90096 MW; DIAD099862528D42 CRC64;

```

```

Query Match 4.7%; Score 143.5; DB 5; Length 782;
Best Local Similarity 21.5%; Pred. No. 0.022;
Matches 112; Conservative 70; Mismatches 175; Indels 165; Gaps 26;

QY 123 SLANYLHKNGET----SLKPEDDF-----TVLSKRGKISRYKDCSMAALTSHLQNS- 171
Db 18 NVADGINVNGDNNGYKTIINDFDDYNYWTPINKKEFLNSYED--KFSSESLENKSS 75
QY 172 -NNSNMNLRTRSKCKKDVMPSPSSSELQESRGLSNFTSTHLLKKEDEGVDDVNFRRVKR 230
Db 76 VDDGINLTDSTSNK-----SSKKGHSRVRSAASAAIIEEDDSKDDMEF----- 122
QY 231 PKGKVTILKGIPIKTKKGRKSCGVSQDSKRESVNCNKADESEPVAKSOLDRTVCI 290
Db 123 -KASPVVKTSTPSGTQTSCLKSS--PSTKSSPSN-----V 158
QY 291 SDAGACGETLSVTSENSLVKKERSLSSGNSFCSEQKTSGLINKFCSAKDSHNEKYED 350
Db 123 -KASPVVKTSTPSGTQTSCLKSS--PSTKSSPSN-----V 158
QY 291 SDAGACGETLSVTSENSLVKKERSLSSGNSFCSEQKTSGLINKFCSAKDSHNEKYED 350
Db 159 KSASPHGES---NSSEESTTKSKRSAS-----VAGIV-----GADEEAPPAPKN 200
QY 351 TFLESEETI-GTKVEVVERKEHLHT-----DILKRGs-----EMDNCSPTRKDFGTG 395
Db 201 TLTPLEELYPTNVNLNFKYSLNNMEENINILKNEGDLVAQKEEFEDENMEKAKODKK 260
QY 396 --EKI-----FQEDTIPRTOIERRTKTSLYFSKYKNEALSPPRRKAFKKWTPPRS 443
Db 261 ALEKIGKESDEAPFMFSSENKFLNQVKNVAGSFRFSFK--LNP-----FKK----- 307
QY 444 PFNLVQETLFDHPWKLLIATIFLNTSGKMAIPVLWKFLKYPSPA-----488
Db 308 -DEVIEKT---EVSKKTFSGIGFNLTEKEAKVLGVGVGYQYETPTMLYNCNPNNSNLFDTI 363
QY 489 -----EVARTADWRDVSSELLKPLGLYDLRAK--TIVKFSDEYLTQWKYPYIEL 534
Db 364 ESLOGRIIDIKKESMISTTFEQOKECLEKNGMVGDLDELNDTQCKFGT-----410
QY 535 HGIGKYGNDSYRIF-CVNEWKQVHP-----EDHKLKNYH 567
Db 411 -CIGSFGHEHLRLRYEFENDLFRPHNPIDYLTADGYKLOKNIH 451

RESULT 10
Q26007

```

```

QY 489 -----EVARTADWRDVSSELLKPLGLYDLRAK--TIVKFSDEYLTQWKYPYIEL 534
Db 364 ESLOGRIIDIKKESMISTTFEQOKECLEKNGMVGDLDELNDTQCKFGT-----410
QY 535 HGIGKYGNDSYRIF-CVNEWKQVHP-----EDHKLKNYH 567
Db 411 -CIGSFGHEHLRLRYEFENDLFRPHNPIDYLTADGYKLOKNIH 451

RESULT 9
Q25730
ID Q25730 PRELIMINARY; PRT: 782 AA.
AC Q25730;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE RHOPTRY ASSOCIATED PROTEIN-1.
GN RAP-1.
OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FC27;
RA Shi Y., Lal A.A.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20985; AAA63681.1;
SQ SEQUENCE 782 AA; 90023 MW; F69E26A2A564C8EA CRC64;

```

```

Query Match 4.6%; Score 141.5; DB 5; Length 782;
Best Local Similarity 21.1%; Pred. No. 0.03;
Matches 110; Conservative 71; Mismatches 176; Indels 165; Gaps 25;

QY 123 SLANYLHKNGET----SLKPEDDF-----TVLSKRGKISRYKDCSMAALTSHLQNS- 171
Db 18 NVADGINVNGDNNGYKTIINDFDDYNYWTPINKKEFLNSYED--EFSSESLENKSS 75
QY 172 -NNSNMNLRTRSKCKKDVMPSPSSSELQESRGLSNFTSTHLLKKEDEGVDDVNFRRVKR 230
Db 76 VDDGINLTDSTSNK-----SSKKGHSRVRSAASAAIIEEDDSKDDMEF----- 122
QY 231 PKGKVTILKGIPIKTKKGRKSCGVSQDSKRESVNCNKADESEPVAKSOLDRTVCI 290
Db 123 -KASPVVKTSTPSGTQTSCLKSS--PSTKSSPSN-----V 158
QY 291 SDAGACGETLSVTSENSLVKKERSLSSGNSFCSEQKTSGLINKFCSAKDSHNEKYED 350
Db 159 KSASPHGES---NSSEESTTKSKRSAS-----VAGIV-----GADEEAPPAPKN 200
QY 351 TFLESEETI-GTKVEVVERKEHLHT-----DILKRGs-----EMDNCSPTRKDFGTG 395
Db 201 TLTPLEELYPTNVNLNFKYSLNNMEENINILKNEGDLVAQKEEFEDENMEKAKODKK 260
QY 396 --EKI-----FQEDTIPRTOIERRTKTSLYFSKYKNEALSPPRRKAFKKWTPPRS 443
Db 261 ALEKIGKESDEAPFMFSSENKFLNQVKNVAGSFRFSFK--LNP-----FKK----- 307
QY 444 PFNLVQETLFDHPWKLLIATIFLNTSGKMAIPVLWKFLKYPSPA-----488
Db 308 -DEVIEKT---EVSKKTFSGIGFNLTEKEAKVLGVGVGYQYETPTMLYNCNPNNSNLFDTI 363
QY 489 -----EVARTADWRDVSSELLKPLGLYDLRAK--TIVKFSDEYLTQWKYPYIEL 534
Db 364 ESLOGRIIDIKKESMISTTFEQOKECLEKNGMVGDLDELNDTQCKFGT-----410
QY 535 HGIGKYGNDSYRIF-CVNEWKQVHP-----EDHKLKNYH 567
Db 411 -CIGSFGHEHLRLRYEFENDLFRPHNPIDYLTADGYKLOKNIH 451

RESULT 10
Q26007

```

```

ID Q26007 PRELIMINARY; PRT; 782 AA.
AC Q26007;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE RHOPTRY ASSOCIATED PROTEIN 1 (ROPI) PRECURSOR (FRAGMENT).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RA "Efficacy of vaccines containing Rhoptyr-Associated Proteins RAP1 and
RT RAP2 of Plasmodium falciparum in Saimiri boliviensis monkeys.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF205283; AAF23404.1; -
FT NON_TER 782
SQ SEQUENCE 782 AA: 90080 MW; 2F2E240D40A4C902 CRC64;

Query Match 4.6%; Score 141.5; DB 5; Length 782;
Best Local Similarity 21.1%; Pred. No. 0.03;
Matches 110; Conservative 71; Mismatches 176; Indels 165; Gaps 25;

QY 123 SLANYLHNKET-----SLKPEDFP-----TVLSKRGIKSRKYKDCSMAALTSHLQNS- 171
Db 18 NVADGINVNGNNYKTIINDFDDYNYWTPINKEFLNSYED--EFSESELENKSS 75
QY 172 -NNSNNLRTSRKCKKDVMPSPSSSELSQESGLSNFTSTHLLKDEGVDDVNFRRKVRK 230
Db 76 VDDGINLTDTSNKN-----SSKKGHSRVSASAAAILEDDSKDDMEF----- 122
QY 231 PKGKVTILKGIPIKTKGCRKSCGFVQSDSKRESCVCKADAESEPVAKQSOLDRTVCI 290
Db 123 -KASPSVVKTSPTSGTQTSGLKSS--PSSTKSSPSN-----V 158
QY 291 SDAGAGETLVTSENSLVKKERSLSSGSNFCSEQKTSIINKFCSAKDSEHNEKYED 350
Db 159 KSAPHGSE---NSSESTTKSKRSAS-----VAGIV-----GADEEAPPAPKN 200
QY 351 TPLESEET-GRKVEVVERKEHLHT-----DILKRG-----EMDNCSPTRKDFGT 395
Db 201 TLTPLEELYPTNVNLFYKSYLNMMENINILKNEGDLVAQKEEFYDENMEKAKQDKK 260
QY 396 --EKI-----FQEDTTPRQIERRKTSLYFSSKYNEALSPPRRKAFKWTTPRS 443
Db 261 ALEKIGKOSDEPPMFSENKFLNOKVKNVAGSFRRFSK--LNP-----FKK----- 307
QY 444 PENLVQETLFDHPWKLLIATIFLNRSTSGKMAIPVLWKFLKYPASA----- 488
Db 308 -DEVIEKT---EVSKKTSFGICFNLTKEAKVLGVGVGYQYPTMYLNCNPNNSLFDTI 363
QY 489 -----EVARTADWRDVSLELKPGLYDLRAK-TIVKFSDEYLTQWKYPIEL 534
Db 364 ESLOGRIIDIKKRESMISTTFEQQKCECLKNMGVLDLNDTQCKFGT----- 410
QY 535 HGICKYGNDSYRIF-CVNEWKQVHP-----EDHKLKHYH 567
Db 411 -CIGSGFGEHLRLYEFENDLRFKHPNIDYLTLDAGYKLOKNH 451

RESULT 11
ID Q90430 PRELIMINARY; PRT; 782 AA.
AC Q90430;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)

```

```

DE RHOPTRY ASSOCIATED PROTEIN 1 (FRAGMENT).
GN RAP1
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RA "Efficacy of vaccines containing Rhoptyr-Associated Proteins RAP1 and
RT RAP2 of Plasmodium falciparum in Saimiri boliviensis monkeys.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF205283; AAF23404.1; -
FT NON_TER 782
SQ SEQUENCE 782 AA: 90131 MW; A0FCB64529C34E59 CRC64;

Query Match 4.6%; Score 140.5; DB 5; Length 782;
Best Local Similarity 21.5%; Pred. No. 0.036;
Matches 112; Conservative 70; Mismatches 175; Indels 165; Gaps 26;

QY 123 SLANYLHNKET-----SLKPEDFP-----TVLSKRGIKSRKYKDCSMAALTSHLQNS- 171
Db 18 NVADGINVNGNNYKTIINDFDDYNYWTPINKEFLNSYED--EFSESELENKSS 75
QY 172 -NNSNNLRTSRKCKKDVMPSPSSSELSQESGLSNFTSTHLLKDEGVDDVNFRRKVRK 230
Db 76 VDDGINLTDTSNKN-----SSKKGHSRVSASAAAILEDDSKDDMEF----- 122
QY 231 PKGKVTILKGIPIKTKGCRKSCGFVQSDSKRESCVCKADAESEPVAKQSOLDRTVCI 290
Db 123 -KASPSVVKTSPTSGTQTSGLK-----SSPSSTKSSPSNV-- 158
QY 291 SDAGAGETLVTSENSLVKKERSLSSGSNFCSEQKTSIINKFCSAKDSEHNEKYED 350
Db 159 KSAPHGSE---NSSESTTKSKRSAS-----VAGIV-----GADEEAPPAPKN 200
QY 351 TPLESEET-GRKVEVVERKEHLHT-----DILKRG-----EMDNCSPTRKDFGT 395
Db 201 TLTPLEELYPTNVNLFYKSYLNMMENINILKNEGDLVAQKEEFYDENMEKAKQDKK 260
QY 396 --EKI-----FQEDTTPRQIERRKTSLYFSSKYNEALSPPRRKAFKWTTPRS 443
Db 261 ALEKIGKOSDEPPMFSENKFLNOKVKNVAGSFRRFSK--LNP-----FKK----- 307
QY 444 PENLVQETLFDHPWKLLIATIFLNRSTSGKMAIPVLWKFLKYPASA----- 488
Db 308 -DEVIEKT---EVSKKTSFGICFNLTKEAKVLGVGVGYQYPTMYLNCNPNNSLFDTI 363
QY 489 -----EVARTADWRDVSLELKPGLYDLRAK-TIVKFSDEYLTQWKYPIEL 534
Db 364 ESLOGRIIDIKKRESMISTTFEQQKCECLKNMGVLDLNDTQCKFGT----- 410
QY 535 HGICKYGNDSYRIF-CVNEWKQVHP-----EDHKLKHYH 567
Db 411 -CIGSGFGEHLRLYEFENDLRFKHPNIDYLTLDAGYKLOKNH 451

RESULT 12
ID Q90414 PRELIMINARY; PRT; 782 AA.
AC Q90414;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE RHOPTRY ASSOCIATED PROTEIN-1.
GN RAP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RA "Efficacy of vaccines containing Rhoptyr-Associated Proteins RAP1 and
RT RAP2 of Plasmodium falciparum in Saimiri boliviensis monkeys.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF205283; AAF23404.1; -
FT NON_TER 782
SQ SEQUENCE 782 AA: 90131 MW; A0FCB64529C34E59 CRC64;

```



```
QY 285 DRTVCISDAGACGETLSVTSEENSLVKKERSISGNSFCSEBOKTSGIINKFCSAKDSEH 344
Db 155 PSNV--KTASPHGES--NSSESTSKSKRSAS-----VSGIV-----GADEEV 194
QY 345 NEKYEDTELESEET-GTKVEVVERKEHLHT-----DILKRGSEM-----DNNCSPT 389
Db 195 PPAPKNTLTPLEELYPTNVLNFKYSLNMEENIILKNEGOLVAQKEEFYDENMEKA 254
QY 390 R--KDTGEKI-----FOEDTIPTOIERRKTSLSYFSKYNKEALSPPRKAFFK 437
Db 255 KEEKKALEKIGKETDEEPFETDDKFLNOKVKNVAGFSRFTSK--LNP-----PKK 307
QY 438 WTPRSPFNVLQFTHDHPKLLIATIFLNRTSGKMAIPVLNKFLEKYP-----486
Db 308 -----DEVIEKT-----EVSKTYSIGFNVDNKEAKILGVGATYOEPETMYLNCNPNNS 357
QY 487 -----SAEVARTAD-----WRDVSSELLPLGLYDLRAK--TIVKFSDEVLTKQW 528
Db 358 HLPDTISSLQGRULIDIKKRENMISTTFEQOKECLNMGMVLDLELKDTECKFGT-----410
QY 529 KYPIELHGKGVGNDYSRYP-CVNEKWQVHP-----EDHKLNYH 567
Db 411 -----CIGSGEHLRLYEFENDLFKPHPNIDYLTADGYKLOKNH 451

RESULT 15
Q9U0G6 PRELIMINARY; PRT: 2209 AA.
AC Q9U0G6;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE VAR, MAL4P2.56.
GN VAR.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=3D7.
RA Devlin K., Bowman S., Churcher C., Harris B., Harris D., Lawson D.,
RA Quail M., Barrell B.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL035475; CAB62897.1;
SQ SEQUENCE 2209 AA; 250779 MW; 7A349F6FEEF9F7CB CRC64;
```

```
Query Match 4.5%; Score 137.5; DB 5; Length 2209;
Best Local Similarity 18.8%; Pred. No. 0.24;
Matches 136; Conservative 97; Mismatches 226; Indels 263; Gaps 37;

QY 8 SLSLGRGAAPTPTSERLVPDPNDLRKEDVAME-----LERVGED---BEQM 53
Db 1176 SVTIGGAEGLQST-----DSKDAARGEXTPLDPSFKRPYPFRYLEEWGQNFCKERRK 1230
QY 54 MIKR--SSECNPLLOPIASQFATAGTECRKSPV-----CG-----W- 90
Db 1231 RLKDIKYEORG--DENI--TRYGSGYGEDCKNNLPENPSTFKDLEYPTCAKYCRFYKKWI 1286
QY 91 -----ERVVKORLFGKTAGDFVYFISPOGLKFRSKSSLYNLHK-----NGE 133
Db 1287 NTKKTEYKQEKIYVQOQKDATSDNGKNYDSCDG-KLKQVASTESFLEKLVQCKDNGE 1345
QY 134 TSLK---PEDFDFTVLSKRGIKRYKDCSMAALTSHLQNSNNSNWLRTSKCKKD---187
Db 1346 GTIKFNGGQTFOHT-----EDCKSCS-----KFRKICDNDKCS 1378
QY 188 -----VFMPPSSSSELOESRLSNFTSTHLLKEDGV-----DDVN-----FRKVR 229
Db 1379 GGNTRKVKDGKTPIDAKEIANNINSPOEVTMLVSDNGATGFKGDDLKEACEGKGFESIR 1438
QY 230 KPGKVTILKGIPIKTKKCKRSCSGFVQSD-----SKRE 265
```

```
Db 1439 KRQWKCDNVGYVVCCKPKEGRNRETVRGEKNDDKHITIRALVTHWQVFLDDYKKIKHKI 1498
QY 266 SVCNKADAESFVAQKSOLDRTVCISD-----292
Db 1499 SHCTKTDOGS--TCQNKCONKCKVCVGEWIPKKREWQOIKDRFLKQYKNKLDDEDFNLRS 1556
QY 293 -----AGA-----CGETLSVTSEENS-----LVKKERS 316
Db 1557 CLETFVLQIGAAYGEDKFKVIKLSVFDQSCGSAIASSQKNGEYKDAIECMLKLEEK 1616
QY 317 LSSGSNFCSEOKTSG-IINKFCSAKDSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDI 375
Db 1617 -----ANKCKEDHSSGEOTEKECQSPSVEDE--DDTLHEETEVE-----KAPEICKDV 1662
QY 376 LKRGSEMDNN--CSP--TRKDF-----TGEKIFQEDTIPRTOIERRKTSLYFSSKYNKE 425
Db 1663 IKAPTEPEERKACDPAPTTPKETSPATDSGKETNTEPYTFQDQSDPTKTPEKGPKPCKS 1722
QY 426 ALSPPRKAFAKKWTPPRSPFNVOETLFDHPWKLLIATIFLNRTSGKMAIPVLNKFLEKY 485
Db 1723 LPQPPROK--REFTP-----SDWKVMSASAF--PWTVGVAFWALSYVWVWKKK 1766
QY 486 PSAEVARTADKRVSEL-LKPLGLYDLRAKT--LVKESDEYLTQWKYPIELHIGIKGYKN 542
Db 1767 SKPRV-----DLFSVMEIPQNDYGMPTLKSNNRYVPYSSGKYRGKTYLY-VE-----GDSGT 1817
QY 543 DS 544
Db 1818 DS 1819
```

Search completed: January 15, 2002, 14:06:31
Job time: 970 sec

us-09-724-296-38.rspt

Tue Jan 15 14:12:18 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:52:31 ; Search time 74.94 Seconds
(without alignments)
322.222 Million cell updates/sec

Title: US-09-724-296-37
Perfect score: 1662
Sequence: 1 MIFRFGVSNAMSLWDASPA.....ELSSIRGVKRTGGALQWKS 317

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 68: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1650.5	99.3	320	2 S55418	UV-endonuclease ho
2	484.5	29.2	322	2 E83974	hypothetical prote
3	359.5	21.6	599	2 S71134	UV-endonuclease -
4	345	20.8	656	2 S55262	UV-endonuclease -
5	293.5	17.7	326	2 C75350	probable UV damage
6	104.5	6.3	315	2 B59093	hypothetical prote
7	101	6.1	283	2 S74063	hypothetical prote
8	100	6.0	823	2 G83905	hypothetical prote
9	98	5.9	331	2 S54263	rep A protein - Ba
10	96.5	5.8	969	2 T38478	RhoGAP/LIM domain
11	95	5.7	775	2 T22200	hypothetical prote
12	94.5	5.7	553	2 T21233	hypothetical prote
13	94.5	5.7	1145	2 A59251	myosin - Acetabula
14	93.5	5.6	805	2 T32377	hypothetical prote
15	93.5	5.6	1036	2 T15665	probable isoleucin
16	93	5.6	674	2 T22733	hypothetical prote
17	92.5	5.6	601	2 T02633	hypothetical prote
18	92	5.5	428	2 G75064	hypothetical prote
19	91.5	5.5	921	2 C81153	type I restriction
20	91.5	5.5	237	2 B82644	5-amino-6-(5-phosp
21	91.5	5.5	522	2 G02522	sorting nexin 1 -
22	91.5	5.5	1409	1 ORFFCP	copla polyprotein
23	91.5	5.5	2334	2 S32920	cell wall-associat
24	91	5.5	558	2 T48150	stress-induced pro
25	91	5.5	852	2 B84001	ribonucleoside-dip
26	91	5.5	1020	2 E86165	hypothetical prote
27	91	5.5	1284	1 WNVZAI	A-type inclusion p
28	90.5	5.4	233	1 R5HG1T	ribosomal protein
29	90.5	5.4	273	2 T15313	hypothetical prote

30 90.5 5.4 448 2 T15589
31 90.5 5.4 496 1 S41192
32 90.5 5.4 625 2 T32739
33 90.5 5.4 629 2 E84610
34 90 5.4 572 1 FOHYJH
35 90 5.4 848 2 C82712
36 90 5.4 2104 2 T38774
37 89.5 5.4 279 2 T16816
38 89.5 5.4 429 2 S29565
39 89.5 5.4 553 2 T24639
40 89.5 5.4 5255 2 T31677
41 89 5.4 489 2 S77357
42 89 5.4 550 2 I64203
43 89 5.4 1033 2 S02168
44 89 5.4 5369 2 T44807
45 88.5 5.3 719 2 A42808

ALIGNMENTS

RESULT 1
S55418
UV-endonuclease homolog ywjd - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 15-Oct-1999
C:Accession: S55418: B70060
R:Glaser, P.; Danchin, A.
Submitted to the EMBL Data Library, May 1995
A:Description: Cloning and sequencing of the Bacillus subtilis chromosomal region fro
A:Reference number: S55414
A:Accession: S55418
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-320 <GUA>
A:Cross-references: EMBL:Z49782; NID:g853752; PIDN:CAA89865.1; PID:g853757
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
tech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivoita, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A:Reference number: A69580; MUID:98044033
A:Accession: B70060
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-320 <KUN>
A:Cross-references: GB:299123; GB:AL009126; NID:g2636240; PIDN:CAB15748.1; PID:el1862
A:Experimental source: strain 168
C:Genetics:
A:Gene: ywjd

Query Match 99.3%; Score 1650.5; DB 2; Length 320;
Best Local Similarity 99.1%; Pred. No. 1.2e-128;
Matches 317; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

Qy 1 MIFRFGVSNAMSLWDASPAKTLTFARYSKLSKTKERKALLTVTKANLNTMRTLHVIIIG 60
|||||
Db 1 MIFRFGVSNAMSLWDASPAKTLTFARYSKLSKTKERKALLTVTKANLNTMRTLHVIIIG 60
|||||
Qy 61 HGIPLYRFSSSIVPLATHPDVMDVTPFQKEFEIGELVKTHQRTSFHPNQFTLTSP 120
|||||
Db 61 HGIPLYRFSSSIVPLATHPDVMDVTPFQKEFEIGELVKTHQRTSFHPNQFTLTSP 120
|||||

121 KESVTKNAVTDMAYHYRMLAMGADRSVINIHIGGAYGNKDTATATQFHNKIQLPQEI 180
Db 121 KESVTKNAVTDMAYHYRMLAMGADRSVINIHIGGAYGNKDTATATQFHNKIQLPQEI 180
Qy 181 ERMTLENDKTYTTEETLQVCEQEDVPFDFHHFYANPDHDLNVALPRMUKTWERIG 240
Db 181 ERMTLENDKTYTTEETLQVCEQEDVPFDFHHFYANPDHDLNVALPRMUKTWERIG 240
Qy 241 LQPKVHLSSPKSEQAIRSHADYVDANF---LLERFRQWGTNIDFMIEAKOKKALLRLMD 297
Db 241 LQPKVHLSSPKSEQAIRSHADYVDANFLLERFRQWGTNIDFMIEAKOKKALLRLMD 300
Qy 298 ELSSIRGVKRIKGGALQWKS 317
Db 301 ELSSIRGVKRIKGGALQWKS 320

RESULT 2
E83974
hypotheical protein BH2597 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C:Accession: E83974
R: Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID: 20263314
A:Accession: E83974
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-322 <STO>
A:Cross-references: GB:AP001516; GB:BA000004; NID:g10175192; PIDN:BAB06316.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH2597

Query Match 29.2%; Score 484.5; DB 2; Length 322;
Best Local Similarity 32.2%; Pred. No. 1.9e-32;
Matches 106; Conservative 67; Mismatches 133; Indels 23; Gaps 6;
Qy 1 MIFRFGVSNAMSLDASPAKTLTARYSKLSKTE---RKEALLTVTKANLNTMRTLHY 57
Db 1 MFIQGYVAMSMELANSPSKTMTATQEKIEDHAGLRK--LERIAKTNLHNLRLK 58
Qy 58 ITHGIGPLRYFSSSIVPLATHP-DVMWDFVTPFOKEFEIGELVKTHLRTSFHPNFTL 116
Db 59 NLAYQLSFRSLSKLVPLVHPHPLTEGWKYEIAEELQAVGEFASQMRIDFHPDHEV 118
Qy 117 FTSPKESVTKNVATDMAYHYRMLAMGADRSVINIHIGGAYGNKDTATATQFHNKIQLP 176
Db 119 LNSEAKEITRRSLQTLVHYKLLKGMEDPRHRCVHLVGGKKKGVEAGLEQFIENTASIP 178
Qy 177 QEIKERMTLENDKTYTTEETLQVCEQEDVPFDFHHFYANPDH-----ADLNVALP 230
Db 179 KSLLSMIMLENDKSTYIDVVLGKLAIPVLDIHH-----HDLVHRSKSLQETWQ 231
Qy 231 RMIKTWERIGLQPKVHLSSPKSEQAIRSHADYVDANFLLERFRQWGT----NIDFMIEAK 286
Db 232 RIVATWEDSPLVPKTHLSSPLSGEDPRHHDYINADRTIAFLHEIGADAVDHLVHMIEAK 291
Qy 287 OKDKALLRLMDLSSIRGVKRIKGGALQW 315
Db 292 KKDALFQMLKDLAEYITVVS KSAVEF 320

RESULT 3
S71134
UV-endonuclease - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 20-Jun-2000
C:Accession: S71134; T39815

R: Takao, M.; Yonemasu, R.; Yamamoto, K.; Yasui, A.
Nucleic Acids Res. 24, 1267-1271, 1996
A:Title: Characterization of a UV endonuclease gene from the fission yeast Schizosacch
A:Reference number: S71134; MUID: 96188660
A:Accession: S71134
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-599 <TA>
A:Cross-references: EMBL:D78571; NID:g1399000; PIDN:BA11415.1; PID:g1399001
A:Experimental source: strain SP972
R: Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Lucas, M.; Gaillardin, C.
submitted to the EMBL Data Library, June 1998
A:Reference number: Z21881
A:Accession: T39815
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-599 <LYN>
A:Cross-references: EMBL:AL023859; PIDN:CAAL19577.1; GSPDB:GN00067; SPDB:SPBC19C7.09c
A:Experimental source: strain 972h; cosmid c19C7
C:Genetics:
A:Gene: UVDE; SPBC19C7.09c
A:Map position: 2

Query Match 21.6%; Score 359.5; DB 2; Length 599;
Best Local Similarity 29.7%; Pred. No. 1e-21;
Matches 83; Conservative 60; Mismatches 111; Indels 25; Gaps 7;
Qy 34 TERKEALLTVTKANLNTM---RTLHYIIGHGIPLYRFSSSIVPLATHPDMWDFVTPP- 89
Db 276 TIORDGLESVKQLGTQNVLDLKLVEWNNINFGIHPMRVSSDLFPFASH--AKYGYTLEFA 333
Qy 90 QKEFREIGELVKTHLRTSFHPNFTLFTSPKESVTKNVATDMAYHYRMLAMGADRS-- 147
Db 334 QSHLEEVGKLANKYNHRLTMHPQYQTQIASPREVVVDASIRDLAYHDEILSRMKLNEQLN 393
Qy 148 --SVINITHIGGAYGNKDTATATQFHNKIQLPQEIKERMTLENDKTYTTEETLQVCEQED 205
Db 394 KDAVLIIHLGGTPEGRKKTLDPRKKNYQRLSDSKARLVLENDVSVQDILLPCQELN 453
Qy 206 VPFVDFHHFYANP---DDHADLNVALPRMKTWERTIGLQPKVHLSSPKSEQA- 256
Db 454 IPLVLDWHHNIIVPGTLREGSLDMLPLTIRETWKGTQKOHYSASADPTAISGMKR 513
Qy 257 RSHADYVDANFLLERFRQWGTNIDFMIEAKOKKALLRL 295
Db 514 RAHSDRV-----DFPPCPDPTMDLMEAKEKEQAVFEL 546

RESULT 4
S55262
UV-endonuclease - Neurospora crassa
C:Species: Neurospora crassa
C:Date: 10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 20-Jun-2000
C:Accession: S55262
R: Yajima, H.; Takao, M.; Yasuhira, S.; Zhao, J.H.; Ishii, C.; Inoue, H.; Yasui, A.
EMBO J. 14, 2393-2399, 1995
A:Title: A eukaryotic gene encoding an endonuclease that specifically repairs DNA dam
A:Reference number: S55262; MUID: 95292980
A:Accession: S55262
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-656 <YAJ>
A:Cross-references: EMBL:D11392; NID:g1526560; PID:g927215
C:Genetics:
A:Gene: uvel

Query Match 20.8%; Score 345; DB 2; Length 656;
Best Local Similarity 32.3%; Pred. No. 1.8e-20;
Matches 87; Conservative 51; Mismatches 105; Indels 26; Gaps 8;
Qy 46 ANLNTMTLHYIIGHGIPLYRFSSSIVPLATHPDMWDFVTPFOKE-FREIGELVKTHQ 104

Db 277 ANARDIVKMLCWNKGIIRFLRSLSEMFPFASHP-VHGKYKLAPFASEVLAEGRAVAELG 335
QY 105 LRTSFHPNQFTLTSFKESVTKNAVDMAYHYRMLFAMGIAD-SSVINIHGGAYGN 160
Db 336 HRLTHPGQFTLGSPRKEVESAIRDLEVHDELLSLLKLPQONRDVAWIIHMGOGFGD 395
QY 161 KDTATAQFONIKOLPOEIKERMTLENDKTYTTEETLOVCEQEDVPFDFHHFYANPD 220
Db 396 KAATLERFKRYARLSQSCNRLVLENDVGWTVHDLDPVCEELNIPMLVDYHHNICFD 455
QY 221 D-ADLNAVLP-RRMKTWRIGLOPKVHLSPKSEQAI-RRSHADYVDANF 267
Db 456 PAHLREGTLDISPKQERANTWKRKIKQKMHYSEP-CDGAVTPRRHRKPRV-510
QY 268 ELERFROMGTNIDFMIEAKOKKALLRLM 296
Db 511 --MTLPPCPDMDLMEAKDKEQAVFELM 537
RESULT 5
C75350
probable UV damage endonuclease - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: C75350
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: C75350
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-326 <WHI>
A:Cross-references: GB:AE002022; GB:AE000513; NID:g6459590; PIDN:AAF11370.1; PID:g645959
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1819
A:Map position: 1

Query Match 17.7%; Score 293.5; DB 2; Length 326;
Best Local Similarity 30.1%; Pred. No. 1.1e-16;
Matches 88; Conservative 51; Mismatches 120; Indels 33; Gaps 11;
QY 21 KTLTFARYSKLSKTERKEALLTVTKANLRMTLHYIIGHGIPLYRFSSSIPLATHPD 80
Db 46 RTVTLRYALSPAERAKLLDLYSSNIKTLRGAADYCAAHDIRYLSLSPML--D 102
QY 81 VMWD-----FVTPQKREFEIGELVKTHQRTSFHPNQFTLTSFKESVTKNAVDMAYH 135
Db 103 LAGDDTGAVALTHLAPQLLEAGHAFTDAGVRLMLHPHQPFTVLNSDRPEVRESSVRAMSAH 162
QY 136 YRMLEAMGIADRSVINIHGGAYGNKDTATAQFONIKOLPOEIKERMTLENDKTYTTE 195
Db 163 ARVMDGGLA-RTWNLL--LHGCKGRGAELALIPDLDPVRLGLENDERAYSPA 219
QY 196 ETLOVCEQEDVPFDFHHFYANP--DDHADLNAV--LPRMKTWRIGLOPK-----VHL 247
Db 220 ELLPICEATGTPVFDAAHHVHDKLPQEDPSVREWVLRARATW-----OPPEWQVVHL 274
QY 248 SSPKSEQAIRSHADYVDANFELERFROMGTNIDFM-IEAKOKKAL--LRLM 296
Db 275 SNGIEGPQDRH-----SHLIADFPISAYADVPWIEVEAKGKEATAALRLM 320

RESULT 6
B59093
hypothetical protein px01-18 - Bacillus anthracis virulence plasmid px01
C:Species: Bacillus anthracis
C:Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 11-May-2000

C:Accession: B59093
R:Okimaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koch
J. Bacteriol. 181, 6509-6515, 1999
A:Title: Sequence and organization of px01, the large Bacillus anthracis plasmid harb
A:Reference number: A59091; MUID:99445483
A:Accession: B59093
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-315 <OKI>
A:Cross-references: GB:AF065404; NID:g4894216; PIDN:AAD32322.1; PID:g4894234
A:Experimental source: strain Sterne
A:Note: similar to integrase-recombinase protein (311 aa), MTH893; Methanobacterium t
C:Genetics:
A:Gene: px01-18
A:Genome: plasmid

Query Match 6.3%; Score 104.5; DB 2; Length 315;
Best Local Similarity 22.8%; Pred. No. 0.42;
Matches 65; Conservative 41; Mismatches 114; Indels 65; Gaps 10;
QY 32 SKTERKEALLTVTKANLRMTLHYIIGHGIPLYRFSSSIPLATHPDVMMDFVTPFOK 91
Db 67 SELEHKRYKSQKTRHRIWVVKLHTELGIVSPLDGIHLSPDQSLSDS--DFITELEE 124
QY 92 EPREIGELVKTHQRTSFHPNQFTLTSFKESVTKNAVDMAYHYRMLFAMGIADRSVIN 151
Db 125 -----TLKQTVLST-----KGTERRQAKYRPL-----IMDRNACI 155
QY 152 IHIGGAYG-----NKDTATAQFONIKOLPOEIK-----ERMTLENDKTY-----TTEE 196
Db 156 INLIYVGLSLOELVSLNMSHIQFARNILIVPGENKLTFRSVSLTWEDTQQLYKYVTIPE 215
QY 197 TLQVCEQEDVPF--VFDFHH-----FYANPDHADLNAVLPKMTWRIGLOPKVHLSS 249
Db 216 PVRPOQHTNNPLVAFDFNRGTFRWYENDAPKALSEVAIOKMIR-----LEV 263
QY 250 PKSEQAIRSHADYVDANFELERFROMGTNIDFMIEAKOKKALLR 294
Db 264 KRAELDRISAQOMRNTFILRLIROGVTEKDLVSMGFKTKISLK 308

RESULT 7
S74063
hypothetical protein c0132 - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 09-Oct-1997 #sequence_revision 24-Oct-1997 #text_change 20-Jun-2000
C:Accession: S74063
R:Sensen, C.W.; Klenk, H.P.; Singh, R.K.; Allard, G.; Chan, C.C.Y.; Liu, Q.Y.; Penny,
Mol. Microbiol. 22, 175-191, 1996
A:Title: Organizational characteristics and information content of an archaeal genome
A:Reference number: S73076; MUID:97055432
A:Accession: S74063
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-283 <SEN>
A:Cross-references: EMBL:Y08256; NID:g1707679; PIDN:CAA69434.1; PID:g1707762
A:Experimental source: strain P2
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, September 199
C:Genetics:
A:Start codon: TTG
C:Superfamily: Methanobacterium thermoautotrophicum endonuclease IV

Query Match 6.1%; Score 101; DB 2; Length 283;
Best Local Similarity 21.6%; Pred. No. 0.7;
Matches 52; Conservative 48; Mismatches 87; Indels 54; Gaps 13;
QY 95 EIGELVKTHQRTSPH-PNQFTLTSFKESV--TKNAVDMAYHYRMLFAMGIADRSVIN 151
Db 59 ETQGVARELGVRLSVIAPYFINLCSEKEKIEASKORILDTADRAELMGADAIA-----112
QY 152 IHIGGAYGNKDTATAQFONIKOLPOEIKERMTLENDKTYTTEETL 198

Db 113 IHI--AFYGMK-PEECYQVKEGLEEVIDKAREMGINVAFGVETMAKETAGTGLDEVI 169
 QY 199 QVCEQED--VPVFDHFHYFANPDHDLNVALPRMKTWERTIGLQPKVHLSPPKSOAI 256
 Db 170 SISKELGVIPYI-DWAHTFARQGEIDYKIIDRLIK--ELGL--THINS-HFESLV 221
 QY 257 RSHADYVD-----ANFLERFRQWCTNIDFMIEAKOKDKALLRMDLSS 301
 Db 222 YRGGYVDEHIPIDANAPPEFLAKELLK-----DISITLICESPELERDAKMKVELR 277
 QY 302 I 302
 Db 278 L 278

RESULT 8

G83905
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
 R: Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
 A:Reference number: A83650; MUID:20263314
 A:Accession: G83905
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-823 <STO>
 A:Cross-references: GB:AP001514; GB:BA000004; NID:gl0174613; PIDN:BA05766.1; GSPDB:GN00
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH2047

Query Match 6.0%; Score 100; DB 2; Length 823;
 Best Local Similarity 23.18; Pred. No. 4.1;
 Matches 75; Conservative 44; Mismatches 101; Indels 104; Gaps 18;

QY 30 KLSKTERK-----FALLTVTKANLRNMTLHYIIGHGIPLRFSSSIVPLATHPDV-MW 83
 Db 332 KRLLEKKRLILIEPLISITEPDSNLERS--WILK-----IKDVEHPDLSLW 377
 QY 84 DFVFPQKEFEIGELVKTHQRTSFH--PNOFTLTPSPKESVTKNVATDMAYHMLEA 141
 Db 378 QERLEWLVKQVAYPYMTLGOKQLFKLESTFTWTGEQP-----RRLG 423
 QY 142 MGIADRSVI-----NIHIGGAYGN-----KDTATAQFH--ONIKO---LPQEI 180
 Db 424 EYADRNLLEYECHPLNNIKIGSIKHLVKQDYPKWLSCAKHGEORRQEQALAEIF 483
 QY 181 ERMTLENDKTYTTEETLQVCEQEDVPFDFHFFHYANPDHDLNVALPRMKTWERTIG 240
 Db 484 KMW-----YPNEDS-----VPFLKFVHDLNHPDVH-----TWKRW 515
 QY 241 LQPKVHL-----SSPKSEQAIRSHADYVDANFLERFRQWGTNIDFMIEAKOKDKA 291
 Db 516 QSIKTEIEATHEVASNPKS--VHLSLDYQD-----FDRDMAWLSPDLMI-AKKDDDS 568
 QY 292 LLRLMDELSSIRGVKRGGGALOW 315
 Db 569 YQVILGE---IHDPTIMVWGVALQF 589

RESULT 9

S54263
 rep A protein - Bacteroides fragilis
 C:Species: Bacteroides fragilis
 C:Date: 08-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 08-Oct-1999
 A:Accession: S54263
 R: Haggoud, A.; Trinh, S.; Mohieddine, M.; Reysset, G.
 submitted to the EMBL Data Library, April 1995

A:Description: Genetic analysis of the minimal replicon of plasmid pIP417 and compari
 A:Reference number: S54263
 A:Accession: S54263
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-331 <HAG>
 A:Cross-references: EMBL:X86702; NID:g804960; PIDN:CAA60390.1; PID:g804961

Query Match 5.9%; Score 98; DB 2; Length 331;
 Best Local Similarity 19.9%; Pred. No. 1.6;
 Matches 52; Conservative 42; Mismatches 85; Indels 82; Gaps 12;

QY 63 IPLYRESSIVPLATHPDVMDVTPFOKEFEIGELVKTHQRTSFHFNQFTLTSPKE 122
 Db 109 LPYIRKRSSVISFOVHPRV-WCILNFSKGYRKY-ELKAAMFKSQFSMRFYELLSNQK- 165
 QY 123 SVTKNAVTDIMAYHMLEAM-GIADRSVINIHIGGAYGNKDTATAQFHONIKOLPQEIKE 181
 Db 166 -----TPLYSIDQLKEMFCVADK-----YKRVNDEMCKVVE 197
 QY 182 RMTLENDKTYTTEETLQVCEQEDVPFDFH-----HFY-ANPDHDLNVAL 229
 Db 198 AAKKELDEVS-----PYTFEFTPLKSGRKITAKEYPYVQPEHRDPL-- 240
 QY 230 PRMIKTWERIGLQPKVHLS---SPKSEQAIRSHADYVDANFLERFRQWGTNIDFMIEAK 286
 Db 241 -----EKHDLQKQVLSWSLSSESVRSVLKNSIGYSD-----KEIKNNLDLFISAQ 285
 QY 287 QKDKALLRLMDELSSTIRGVKR 307
 Db 286 ---SLPDLILGELAILKGRSR 303

RESULT 10

RHO478
 RHO478/LIM domain protein - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: T38478
 R: Brown, D.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, June 1997
 A:Reference number: Z1796
 A:Accession: T38478
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-969 <BRO>
 A:Cross-references: EMBL:Z97210; PIDN:CAB10138.1; GSPDB:GN00066; SPDB:SPAC29A4.11
 A:Experimental source: strain 972h; cosmid c29A4
 C:Genetics:
 A:Gene: SPDB:SPAC29A4.11
 A:Map position: 1

Query Match 5.8%; Score 96.5; DB 2; Length 969;
 Best Local Similarity 18.5%; Pred. No. 10;
 Matches 60; Conservative 52; Mismatches 99; Indels 113; Gaps 15;

QY 17 ASPAKTLFARYSKISKTERKEALLTVTKANLRNMTLHYIIGHCIPL---YRFSSTIV 73
 Db 321 ASPKYT-----MSLTDRAPIV-----MNGHMRSLHNATSPFRPFSYRSSDTHS 366
 QY 74 PLATHPDVMDVTPFOKEFEIGELVKTHQRTSFHFNQFTLT---SPKESVTKNAV 130
 Db 367 PTRSPN-----VOTHK-KTSSQPSDLSLSPQVLSPPKPN 406
 QY 131 DMAYHMLEAMGIADRSVINIHIGG-----AYGNKDTATAQFH---QNI 172
 Db 407 GGHKSFRRSHLSLSETSQOTLVPSLGNGEYHLPTNDHSSTPAQSERDSIDVELREOLENL 466
 QY 173 KOLPQEIKEKMT-----LENDKTYTTEETLQVCEQEDVPFDFHFFHYANPDH 223
 Db 467 TALTKKLSRSLSSSTFDSKFIREDKOTVRSKLEICEK-----FFSFADVTDPT 518

SECRET/NOFORN

QY 273 RQWGTNIDFMIEAKOKKALLRLMDELSSIRGV 305
 Db 544 EGG-----KNAKGIFPLIDEACKMPNV 567

RESULT 14

T32377
 hypothetical protein K10F12.5 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 01-Dec-2000
 C:Accession: T32377; T22150; T16200; T22352
 R:Wohlmann, P.; Beck, C.
 submitted to the EMBL Data Library, September 1997
 A:Description: The sequence of C. elegans cosmid K10F12.
 A:Reference number: Z21157
 A:Accession: T32377
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 1-805 <WOH>
 A:Cross-references: EMBL:AF025462; PIDN:AAB71003.1; GSPDB:GN00021; CESP:K10F12.5
 A:Experimental source: strain Bristol N2; clone K10F12
 R:Kershaw, J.
 submitted to the EMBL Data Library, September 1996
 A:Reference number: Z19523
 A:Accession: T22150
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 1-120, 'F', 122-280, 'P', 282-551, 'T', 553-577, 'A', 579-805 <WIL>
 A:Cross-references: EMBL:Z79755; PIDN:CAB02112.1; GSPDB:GN00019; CESP:F43G9.7
 A:Experimental source: clone F43G9
 R:Le, T.
 submitted to the EMBL Data Library, November 1995
 A:Description: The sequence of C. elegans cosmid F28E10.
 A:Reference number: Z18476
 A:Accession: T16200
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 253-551, 'T', 553-805 <LET>
 A:Cross-references: EMBL:U040801; NID:g1065958; PID:g1065959; PIDN:AAA81500.1; CESP:F28E10
 R:Mathews, L.
 submitted to the EMBL Data Library, August 1996
 A:Reference number: Z19552
 A:Accession: T22352
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 253-551, 'T', 553-697, 'L', 699-805 <WT2>
 A:Cross-references: EMBL:Z79599; PIDN:CAB01870.1; GSPDB:GN00028; CESP:F47C8.1
 A:Experimental source: clone F47C8
 C:Genetics:
 A:Gene: CESP:K10F12.5; CESP:F43G9.7; CESP:F28E10.3; CESP:F47C8.1
 A:Map position: 3

Query Match 5.6%; Score 93.5; DB 2; Length 805;
 Best Local Similarity 20.8%; Pred. No. 14;
 Matches 56; Conservative 41; Mismatches 123; Indels 49; Gaps 10;

QY 60 GHGIPLYRFSSSIPLATHPDVMDVTPQKEFREIGELV---KTHQLTSFHPNOFTL 116
 Db 531 GRNLTNLRFADDIVLIANHPNIASIKLOELVQKSEVGLEINTGKTKVLNRNRFADPSKVY 590

QY 117 FTSPKESVTKNAVTDMAHYHRLMEAMGIAIDRSVINIHIGGAYGNKDTATAQFHNKQLP 176
 Db 591 FGSPSPPTOLDVDVYILGRQINAQ---NNLMPEIH-----RRRAAWAFN-GIKNTT 641

QY 177 QFIKERMLEN-----DKTYTTEE---TLQVCEQEDVPFVDFHHFYANPDDHADLN 226
 Db 642 DSITDKKIRANLFDISIVLPALTYGSEAWTFKALSERVRI-----THASLE 687

QY 227 VALPRMKTWERIGLQPKVHLSFKSPQAIRSHADYVANDANFLERFROWGTNIDEMIEAK 286
 Db 688 RRLVGIITLQOR---ERDLHREDIRTNLSYRD-----PLNFVKRKLGNAGHV-----AR 734

QY 287 QKDKALLRLMDELSSIRGVKRIIGGALQW 315
 Db 735 RKDGRWTTLTMTWRPYGKWKRPVGRPPMRW 763

RESULT 15

F71565
 probable isoleucine--tRNA ligase (EC 6.1.1.5) - Chlamydia trachomatis (serotype D, st
 C:Species: Chlamydia trachomatis
 C>Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 24-Nov-1999
 C:Accession: F71565
 R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
 Science 282, 754-759, 1998
 A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
 A:Reference number: A71570; MUID:99000809
 A:Accession: F71565
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1036 <ARN>
 A:Cross-references: GB:AE001277; GB:AE001273; NID:g3328408; PIDN:AAC67609.1; PID:g332
 A:Experimental source: serotype D, strain UW-3/Cx
 C:Genetics:
 A:Gene: files
 C:Superfamily: isoleucine--tRNA ligase
 C:Keywords: ligase

Query Match 5.6%; Score 93.5; DB 2; Length 1036;
 Best Local Similarity 19.7%; Pred. No. 20;
 Matches 57; Conservative 42; Mismatches 104; Indels 87; Gaps 13;

QY 71 SIPLATHPDVMDVTPQKEFREIGELVKTQLTSFHP 111
 Db 229 SNMALAVHVELTYVRKDKESGDEYILGQESLPWFDPRESYEWIGLSGKSLVGQSYEP 288

QY 112 NQFTLTSPKESVTKNAVTDMAHYHRLMEAMGIAIDRSVIN-----IHIGGAYGNKDTA 164
 Db 289 -LPPYFQDKKE-----LEAFRLPADFTIESESEGTGIVHMAPAFGEADFF 331

QY 165 TAOFH-----QNIKQLPQEIKERMTLENDKTY---TTEETL 198
 Db 332 ACQEHNVPLVCPVDNOCGYTAEVKDFVG EYIKSADKGIARR--LKNENKLFYQGTVRHRY 389

QY 199 QVCEQEDVPFVDFHHFYANPDDHADLNVALPRMKTWERIGLQPKVHLSFKSPQAIRS 258
 Db 390 PFCWRDTSPLIYKAVNSW-----FVAVERKVKSKMLKANESIHWTPE-HLKQGRFGKWLEG 443

QY 259 HADYVDANELLERFROWGTNIDFMIEAKOKKALLRL--MDELSIRGVK 306
 Db 444 ARDWA-----ISRNYWGTPIPIW---RSDDGELLVIGSIQLEALSQK 485

Search completed: January 15, 2002, 13:52:32
 Job time: 401 sec

us-09-724-296-37.rpr

Tue Jan 15 14:12:14 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 14:07:17 ; Search time 45.07 seconds
(without alignments)
257.882 Million cell updates/sec

Title: US-09-724-296-37

Perfect score: 1662

Sequence: 1 MIFRFGVSNMGLWDASPA.....ELSSINGVKRIGCGALQWKS (317)

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1650.5	99.3	320	YWJD_BACSU	P45864 bacillus su
2	103	6.2	964	CTA2_BACCI	P70873 bacillus ci
3	93.5	5.6	1036	SYI_CHLTR	O84022 chlamydia ci
4	92.5	5.6	501	C4D2_DROME	Q27589 drosophila t
5	91.5	5.5	522	SNX1_HUMAN	Q13596 homo sapien
6	91.5	5.5	1409	COPI_DROME	P04146 drosophila
7	91.5	5.5	2334	WAPA_BACSU	Q07833 bacillus su
8	91	5.5	1284	ATI_COMPLEX	P16602 cowpox viru
9	90.5	5.4	233	RL1_THEMA	P29393 thermotoga
10	90	5.4	572	GAG_IPHA	P04023 hamster int
11	90	5.4	1941	MYH2_HUMAN	Q9UKX2 homo sapien
12	90	5.4	2104	MYS3_SCHPO	O14157 schizosacch
13	89.5	5.4	429	APAA_MACFA	P33621 macaca fasc
14	89.5	5.4	5255	BACA_BACLI	O68006 b bacitraci
15	89	5.4	550	SYD_MYCGE	P47282 mycoplasma
16	89	5.4	1033	TIRI_ECOLI	P10486 escherichia
17	88.5	5.3	582	HEM0_OPSTA	P43090 opsanus tau
18	87.5	5.3	856	VPH1_NEUCR	Q01290 neurospora
19	86.5	5.2	533	KCCD_RAT	P15791 rattus norv
20	86.5	5.2	802	PUR2_YEAST	P07244 s bifunctio
21	86.5	5.2	1259	LIN1_HUMAN	P08547 homo sapien
22	86	5.2	427	RHO_THEMA	P38527 thermotoga
23	86	5.2	583	FOJO_DROME	P54360 drosophila
24	86	5.2	1083	UL52_VZVD	P09270 varicella-z
25	86	5.2	1131	MYPE_CHICK	P16419 gallus gall
26	85	5.1	326	HEMZ_PASMU	P57874 pasteurella
27	85	5.1	601	DR11_MOUSE	Q62431 mus musculu
28	84.5	5.1	382	MATB_NEUCR	P36981 neurospora
29	84.5	5.1	399	RPSP_THEMA	P77994 thermotoga
30	84.5	5.1	508	CPV1_RAT	P22443 rattus norv
31	84.5	5.1	528	MASY_EMENI	P28344 emeritella
32	84.5	5.1	635	SVY_CHLPN	Q927a0 chlamydia p
33	84.5	5.1	1085	CUT7_SCHPO	P24339 schizosacch

How did they know so definitely?

NO PA

34	84.5	5.1	1679	1	YIO9_YEAST	P40457 saccharomyc
35	84.5	5.1	1739	1	CHD2_HUMAN	O14647 homo sapien
36	84	5.1	276	1	COX2_HUMAN	O9y6n1 homo sapien
37	84	5.1	649	1	METX_YEAST	Q04533 saccharomyc
38	84	5.1	698	1	SST2_YEAST	P11972 saccharomyc
39	84	5.1	724	1	AT11_VACCV	P24759 vaccinia vi
40	84	5.1	843	1	YMS1_YEAST	Q05050 saccharomyc
41	83.5	5.0	436	1	EP45_XENLA	Q00387 xenopus lae
42	83.5	5.0	829	1	GUTR_BACSU	P39143 bacillus su
43	83.5	5.0	1251	1	RBP2_PLAYVB	Q00799 plasmodium
44	83	5.0	343	1	METE_THEAC	P57704 thermoplasm
45	83	5.0	944	1	NUF1_YEAST	P32380 saccharomyc

ALIGNMENTS

RESULT	1
YWJD_BACSU	
ID	YWJD_BACSU
AC	P45864
DT	01-NOV-1995 (Rel. 32, Created)
DT	01-NOV-1995 (Rel. 32, Last sequence update)
DT	20-AUG-2001 (Rel. 40, Last annotation update)
DE	HYPOTHETICAL 36.9 KDA PROTEIN IN ACDA 5 REGION.
GN	YWJD.
OS	Bacillus subtilis.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;
OC	Bacillus/Staphylococcus group; Bacillus.
OX	NCBI_TaxID=1423;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=168;
RA	Glaser P., de la Fuente V., Danchin A.;
RL	Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
EMBL	Z49782; CAA89865.1; -
DR	EMBL; Z99123; CAB15748.1; -
DR	Subtilist; BG11309; YWJD.
KW	Hypothetical protein; Complete proteome.
SQ	SEQUENCE 320 AA; 36896 MW; 2FD35A710648E002 CRC64;

Query Match	99.3%;	Score	1650.5;	DB 1;	Length	320;	
Best Local Similarity	99.1%;	Pred. No.	1.1e-127;				
Matches	317;	Conservative	0;	Mismatches	0;	Gaps	1;
Qy	1	MIFRFGVSNMGLWDASPAKTLTFRYSKLSKTERKEALLTVTKANLRNTRTLHYIG	60				
Db	1	MIFRFGVSNMGLWDASPAKTLTFRYSKLSKTERKEALLTVTKANLRNTRTLHYIG	60				
Qy	61	HGIPLYRFSSSIPLATHPVDMDFVTPFOKEPREIGELVKTHOLRTPSPHNOFTLTSP	120				
Db	61	HGIPLYRFSSSIPLATHPVDMDFVTPFOKEPREIGELVKTHOLRTPSPHNOFTLTSP	120				
Qy	121	KESVTKNAVTDMAHYHRLMELAMGADRSVINIHIGAYGNKDTATATQFHNKQLPOEIK	180				
Db	121	KESVTKNAVTDMAHYHRLMELAMGADRSVINIHIGAYGNKDTATATQFHNKQLPOEIK	180				
Qy	181	ERMTLENDKTYTTEETLQVCEQEDVPFVDFHHFYANPDHDLNVALPRMKTWERTG	240				
Db	181	ERMTLENDKTYTTEETLQVCEQEDVPFVDFHHFYANPDHDLNVALPRMKTWERTG	240				
Qy	241	LQPKVHLSSPSQGAIRSHADYDANF---LLERFRQWGTNIDFMIEAKQKDKALLRLMD	297				

```

Db 241 LQPKVHLSKSPKSEQAIRSHADYVDANFLPLLEFRFROWGNTNIDFMIEAKOKDKALLRLMD 300
QY 298 ELSISRGVKGRIKGGALOWKS 317
Db 301 ELSISRGVKGRIKGGALOWKS 320

RESULT 2
CTA2_BACCI STANDARD; PRT; 964 AA.
ID P70873;
AC 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE CYCLOISOMALTOOLIGOSACCHARIDE GLUCANOTRANSFERASE PRECURSOR (EC 2.4.1.-)
DE (CITASE).
OS Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1397;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=U-155;
RA Oguma T., Kurokawa T., Tobe K., Kitao S., Kobayashi M.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC FUNCTION: PRODUCES CYCLOISOMALTOOLIGOSACCHARIDE FROM DEXTRAN.
CC -!- SIMILARITY: BELONGS TO FAMILY 66 OF GLYCOSYL HYDROLASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; D88360; BAA13595.1;
KW Transferase; Glycosyltransferase; Signal.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 964 CYCLOISOMALTOOLIGOSACCHARIDE
FT GLUCANOTRANSFERASE.
FT SEQUENCE 964 AA; 107208 MW; 8849CDC3E2DE9A68 CRC64;

Query Match 6.2%; Score 103; DB 1; Length 964;
Best Local Similarity 22.1%; Pred. No. 1.2;
Matches 55; Conservative 44; Mismatches 84; Indels 66; Gaps 14;

QY 95 EIGELVKTHTLRTSFHPNQFTLTSPKSVTKNAVTDMAHYRMLMANG--IADRSVIN- 151
Db 561 ELGE--DSQMLAHEYPNR-----SKSMRSLKSAKMD---HYNFTAYENLLFDADVDIN 611
QY 152 -----IHIGGAYGNKDTATQAFONIKOLPO-ETKERMILENDKTY-----TTEET- 197
Db 612 DAGKQFINTAGVNTSPDGAANTVHMSKRTPEYNTLHLINLVNDQNWNSNQTAQTN 671
QY 198 ----LQVCEQEDVPFVDFHFHYANPDHDLNVALP-----RMKWTWE 237
Db 672 LATKVIYGAETITGV-----YAAPDHQOGATQSLPFTTGTSSGYSFTSFTVPSLEWXS 726
QY 238 RIGLQPKVHLSPP-----KSEQAIRS-----HADYVDANFLLERFROWGNTNIDFMIEA 285
Db 727 MIYN--KRSTAAPVDNMVYAEATAIKSNVSVNTNHAGYTGSGF--VDQFATVNDGVSVFVHA 783
QY 286 KOKDKALLR 294
Db 784 SSKDDYVLR 792

RESULT 3
SYL_CHLTR
ID SYL_CHLTR STANDARD; PRT; 1036 AA.

```

```

084022;
20-AUG-2001 (Rel. 40, Created)
20-AUG-2001 (Rel. 40, Last sequence update)
20-AUG-2001 (Rel. 40, Last annotation update)
DE ISOLEUCYL-TRNA SYNTHETASE (EC 6.1.1.5) (ISOLEUCINE--TRNA LIGASE)
DE (ILERS).
GN LES OR CT019.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/CX;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
CC -!- CATALYTIC ACTIVITY: ATP + L-ISOLEUCINE + TRNA(ILE) = AMP +
CC PYROPHOSPHATE + L-ISOLEUCYL-TRNA(ILE).
CC -!- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; AF001277; AAC67609.1;
DR HSSP; P56690; 11LE.
DR InterPro; IPR002300; trna-synt_la.
DR InterPro; IPR001412; trna-synt_l.
DR InterPro; IPR002301; trna-synt_ile.
DR Pfam; PF00133; trna-synt_1; 1.
DR PRINTS; PR00984; TRNASYNTHILE.
DR PROSITE; PS00178; AA-TRNA-LIGASE_I; 1.
KW Aminoacyl-TRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Metal-binding; Zinc; Complete proteome.
FT SITE 46 56 "HIGH" REGION.
FT SITE 589 593 "KMSKS" REGION.
FT BINDING 592 592 ATP (BY SIMILARITY).
FT SEQUENCE 1036 AA; 118985 MW; DC4420FD3734EFD CRC64;

Query Match 5.6%; Score 93.5; DB 1; Length 1036;
Best Local Similarity 19.7%; Pred. No. 8;
Matches 57; Conservative 42; Mismatches 104; Indels 87; Gaps 13;

QY 71 STVPLATHPDVMMW-----DFVT-----PFOKEFREIGELVKTHTLRTSFHP 111
Db 229 SNMALAVHPELTYVRIKDKESGEYILQOESLPWFDPRESYEWIGSLGSLVGQSYEP 288
QY 112 NOFTLTSPKESVTKNAVTDMAHYRMLMANGTADRSVIN-----IHIGGAYGNKDTA 164
Db 289 -LFYFQDKKE-----LEAFRIPLADFIESEGTGIVHMAPAFGEADFF 331
QY 165 TAOEH-----QNIKOLPOETKERMILENDKTY---TTEETL 198
Db 332 ACQEHNVPLVCPVDNNOGCYTAEVKDFVGEYIKSADKGIARR--LKNENKLFYQCTVRRHY 389
QY 199 QVCEQEDVPFVDFHFHYANPDHDLNVALPRMKTWTWIRIGLQPKVHLSKSPKSEQAIRS 258
Db 390 PFCWRTDPLIYKAVNSW-----EVAVEKVKSKMLKANESITHWTPE-HIKGREGKWLKG 443
QY 259 HADYVDANFLLERFROWGNTNIDFMIEAKOKDKALLRL--MDELSISRGVK 306

```

Db 444 ARDWA-----ISNRNWTGTPITW---RSDGELLVIGSIQLEALSQK 485

RESULT 4

C4D2_DROME STANDARD; PRt; 501 AA.

AC Q27589; Q27588; Q46053; Q18651; Q18674; Q9W514;

DT 15-DEC-1998 (Rel. 37, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE CYTOCHROME P450 4D2 (EC 1.14.1.1) (CYP1D2).

GN CYP4D2 OR EG:152A3.4 OR CG3466.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidae; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=OREGON-R;

RX MEDLINE=94296569; PubMed=8024706;

RA Frolow M.V., Alatorsev V.E.;

RT "Cluster of cytochrome P450 genes on the X chromosome of Drosophila melanogaster.";

RL DNA Cell Biol. 13:663-668(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=OREGON-R;

RX MEDLINE=20196011; PubMed=10731137;

RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,

RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,

RA Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,

RA Minana B., Kafatos G., Louis C., Siden-Kiamos I., Bolshakov S.,

RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablo B.,

RA Modolell J., Peter A., Schoettler J., Werner M., Mourkioti F.,

RA Beinert N., Dowe G., Schaefer U., Jaekle H., Bucheton A.,

RA Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,

RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,

RA Glover D.M.;

RT "From sequence to chromosome: the tip of the X chromosome of D. melanogaster.";

RL Science 287:2220-2222(2000).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abail J.F., Agbayani A., An H.-J., Andrews-pfannkuch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Fiesler K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwac C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merklov G., Hishina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

RN [4]

RP SEQUENCE OF 4-501 FROM N.A.

RC STRAIN-VARIOUS STRAINS;

RA Phillips K.S., Begun D.J., Aquadro C.F.;

RT "Evidence for non-neutral evolution around the cytochrome p450 gene cluster on the Drosophila melanogaster X chromosome.";

RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: INVOLVED IN THE METABOLISM OF INSECT HORMONES AND IN THE BREAKDOWN OF SYNTHETIC INSECTICIDES (BY SIMILARITY).

CC -!- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH + OXIDIZED FLAVOPROTEIN + H(2)O.

CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (POTENTIAL).

CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; X75955; CAA53568.1; -

CC EMBL; Z23005; CAA80549.1; -

CC EMBL; AL009194; CAA15698.1; -

CC EMBL; AE003423; AAF45741.1; -

CC EMBL; AF017006; AAB71169.1; -

CC EMBL; AF017007; AAB71170.1; -

CC EMBL; AF017008; AAB71171.1; -

CC EMBL; AF017009; AAB71172.1; -

CC EMBL; AF017010; AAB71173.1; -

CC EMBL; AF017011; AAB71174.1; -

CC EMBL; AF017012; AAB71175.1; -

CC EMBL; AF017013; AAB71176.1; -

CC EMBL; AF017014; AAB71177.1; -

CC EMBL; AF017015; AAB71178.1; -

CC EMBL; AF017016; AAB71179.1; -

CC EMBL; AF017017; AAB71180.1; -

CC EMBL; AF017018; AAB71181.1; -

CC FlyBase; FBgn0011576; Cyp4d2.

CC InterPro; IPR001128; Cyt_P450.

CC Pfam; PF00067; p450; 1.

CC PRINTS; PR00385; P450.

CC PRINTS; PR00464; EP450I1.

CC PROSITE; PS00086; CYTOCHROME_P450; 1.

CC Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;

CC Endoplasmic reticulum; Polymorphism.

FT BINDING 449 449 HEME (BY SIMILARITY).

FT BINDING 163 163 K -> M (IN STRAINS CAM-44, CAM-48 AND BERKELEY).

FT CONFLICT 2 25 MISSING (IN REF. 2).

FT CONFLICT 30 30 I -> A (IN REF. 1; CAA80549).

FT CONFLICT 160 160 A -> R (IN REF. 1).

FT CONFLICT 487 501 LRSANGVHGLKPRA -> CGRPTAFILA (IN REF. 1).

FT SQ SEQUENCE 501 AA; 57914 MW; 8F8D98DBF39FD1A2 CRC64;

Query Match 5.6%; Score 92.5; DB 1; Length 501;

Best Local Similarity 19.8%; Pred. No. 3.6;

Matches 55; Conservative 43; Mismatches 91; Indels 89; Gaps 12;

[illegible]


```
DR EMBL: 299124; CAB15959.1; --
DR PIR: S32920; S32920.
DR Subtilisin; BG10797; wpaA.
DR InterPro: IPR003305; CBD_6.
DR Pfam: PF02018; CBD_6; 1.
DR Cell wall; Repeat; Signal; Complete proteome.
DR SIGNAL 1 28
DR CHAIN 29 2334
DR DOMAIN 504 869
DR REPEAT 504 605
DR REPEAT 636 736
DR REPEAT 769 869
DR DOMAIN 1021 2139
DR REPEAT 1021 1040
DR REPEAT 1042 1061
DR REPEAT 1063 1082
DR REPEAT 1083 1102
DR REPEAT 1109 1128
DR REPEAT 1129 1148
DR REPEAT 1150 1169
DR REPEAT 1174 1193
DR REPEAT 1199 1218
DR REPEAT 1219 1238
DR REPEAT 1265 1284
DR REPEAT 1646 1665
DR REPEAT 1667 1686
DR REPEAT 1690 1709
DR REPEAT 1711 1730
DR REPEAT 1732 1751
DR REPEAT 1753 1772
DR REPEAT 1795 1814
DR REPEAT 1820 1839
DR REPEAT 1840 1859
DR REPEAT 1861 1880
DR REPEAT 1887 1906
DR REPEAT 1908 1927
DR REPEAT 1929 1948
DR REPEAT 1969 1982
DR REPEAT 1983 2002
DR REPEAT 2008 2027
DR REPEAT 2028 2047
DR REPEAT 2051 2070
DR REPEAT 2071 2090
DR REPEAT 2093 2112
DR REPEAT 2120 2139
DR SEQUENCE 2334 AA; 258329 MW; B75138CCD278BAA3 CRC64;

Query Match
Best Local Similarity 22.5%; Score 91.5; DB 1; Length 2334;
Matches 63; Conservative 35; Mismatches 115; Indels 67; Gaps 13;

QY 33 KTERKEALLTVTKANLRNMTLHYIGHGIPLYRFSSIVPLATH---PDVMDVFTVPFQ 90
Db 66 KOINKEGETTSR-----TENTKLYEGDV--YKQEVYLDPIHTKETPDADWEDISPEL 118
QY 91 KEFREIGELVKTHQLTSFHPNQFTLFTSPKESVTKNAV-----TDMAYHYRMLEAMGI 144
Db 119 KE--STSKOVETEN-----AILNSDFQKMKNGLYATFEHNDHKVYISAEAKGP 166
QY 145 ADRSVI-----NIHIGGAYGNKDTATQAQFHNKQLPQETIKERMTE--NDD 189
Db 167 NKTSLTPKDSADYKTDNSNEIVYDPVFPFNIDLTFTFEN-----IKEDLVHLQYNGY 219
QY 190 KYTT--TETLQVCQEQDVPFVDFHHFYANFPDDHADLNVALPRMKTWER-----IG 240
Db 220 NTFTTQKTDLOAKQEQEQSIDFS-----DEKGVVFSVPKPFMTDSKLDLSGEVE 271
QY 241 LQPKVHLSPPKSEQAQRSHADYVDANFLERFRQWGTNID 280
Db 272 RSDKVSYLEKNEEGVLLHLT--ADENWLKDPERVYVPSVD 310
```

```
RESULT 8
ATL_COWPX STANDARD; PRT; 1284 AA.
AC P16602;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE A-TYPE INCLUSION PROTEIN (ATI).
GN ATI.
OS Cowpox virus (CPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10243;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CPRO6;
RX MEDLINE=88089536; PubMed=2826668;
RA Funahashi S., Sato T., Shida H.;
RT "Cloning and characterization of the gene encoding the major protein
of the A-type inclusion body of cowpox virus.";
RL J. Gen. Virol. 69:35-47(1988).
RN [2]
RP SEQUENCE OF 1-109 FROM N.A.
RX MEDLINE=88111568; PubMed=2828037;
RA Patel D.D., Pickup D.J.;
RT "Messenger RNAs of a strongly-expressed late gene of cowpox virus
contain 5'-terminal poly(A) sequences.";
RL EMBO J. 6:3787-3794(1987).
CC !- FUNCTION: MATURE VIRUSES ARE OCCLUDED INTO THE ATI, AND IT HAS
BEEN ASSUMED THAT SUCH BODIES PROTECT THE VIRUS DURING
DISSEMINATION FROM ANIMAL TO ANIMAL.
CC !- MISCELLANEOUS: A CHARACTERISTIC FEATURE OF ATI IS THE FORMATION OF
LARGE MASSES WITH NO SURROUNDING MEMBRANES IN THE CYTOPLASM OF
INFECTED CELLS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL: D00319; BAA00222.1; --
CC PIR: JQ0006; WHVZAI.
CC Late protein; Repeat.
KW DOMAIN 611 912
FT REPEAT 611 637 1.
FT REPEAT 638 665 2.
FT REPEAT 666 689 3.
FT REPEAT 690 720 4.
FT REPEAT 721 751 5.
FT REPEAT 752 780 6.
FT REPEAT 781 811 7.
FT REPEAT 812 842 8.
FT REPEAT 843 871 9.
FT REPEAT 872 912 10.
SQ SEQUENCE 1284 AA; 150329 MW; F7904C9E1DE8D012 CRC64;

10 X APPROXIMATE TANDEM REPEATS.

Query Match
Best Local Similarity 19.4%; Score 91; DB 1; Length 1284;
Matches 60; Conservative 50; Mismatches 99; Indels 100; Gaps 15;

QY 56 HYIIGHGIPLY-----RFSSIVPLATHPDVMDVFTVPFQEFREIGELVKTHQLRT 107
Db 190 HYLLVSGSELFINVAGFASFSSPI-----SVDPYIMEALT-----FKACDIHMKSGDLK- 239
QY 108 SFHPNQFTLFTSPKESVTKNAVDTMAYHYRMLEAM---GIADR-----SVINTHIGG 156
Db 240 -----YTYAFTKVKDLFNFTKSDSVYQVYRLHEMSYDGVSEDTDDDDDEFAILLNLSI-- 291
```



```

RESULT 12
MYS3_SCHPO STANDARD; PRT: 2104 AA.
AC 014157: 042730:
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN TYPE II HEAVY CHAIN 2.
GN MYO3 OR MYP2 OR SPAC4A8.05C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98119390; PubMed=9459302;
RA Motegi F., Nakano K., Kitayama C., Yamamoto M., Mabuchi I.;
RT "Identification of a second myosin-II in Schizosaccharomyces pombe:
RT fission yeast Schizosaccharomyces pombe.";
RL FEBS Lett. 420:161-166(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98062346; PubMed=9398685;
RA Bezanilla M., Forsburg S.L., Pollard T.D.;
RT "Identification of a second myosin-II in Schizosaccharomyces pombe:
RT Myo2p is conditionally required for cytokinesis.";
RL Mol. Biol. Cell 8:2693-2705(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Wood V.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: STABILIZES THE F-ACTIN CABLES FORMING THE F-ACTIN RING
CC THAT SURROUNDS THE NUCLEUS DURING INTERPHASE. MAY WORK IN
CC CONJUNCTION WITH MYO2.
CC -!- SUBUNIT: BINDS TO CDC4 AND RLC1.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB007633; BAA24579.1; -
CC EMBL; AF029788; AAC04615.1; -
CC EMBL; Z98762; CAB11475.1; -
CC HSSP; P08799; 1MND.
CC InterPro; IPR001609; myosin_head.
CC Pfam; PF00063; myosin_head; 1.
CC PRINTS; PR00193; MYOSINHEAVY.
CC ProDom; PD000355; myosin_head; 1.
CC SMART; SM00242; MYSC; 1.
CC Myosin; Coiled coil; ATP-binding; Actin-binding; Alkylation.
KW DOMAIN 1 828 MYOSIN HEAD-LIKE.
FT FT DOMAIN 829 2104 COILED COIL (POTENTIAL).
FT FT DOMAIN 646 660 ACTIN-BINDING (POTENTIAL).
FT FT NP_BIND 182 189 ATP (POTENTIAL).
FT FT MOD_RES 686 686 ALKYLATION (BY SIMILARITY).
FT FT CONFLICT 1193 1193 D -> G (IN REF. 2).
FT FT CONFLICT 1304 1304 E -> G (IN REF. 2).
FT FT CONFLICT 1344 1344 E -> K (IN REF. 2).
FT FT CONFLICT 1420 1420 G -> D (IN REF. 2).
FT FT CONFLICT 1420 1420 G -> D (IN REF. 2).
SQ SEQUENCE 2104 AA; 242570 MW; 3A0548594028D258 CRC64;

22 TLTFARYSKLTKERKALLT-----VTKANLRMTMTLHVLIIGHIPLYRFSIVPL 75
   || || || || || || || || || || || || || || || || || || ||
863 TITKERVQLTNDLQEEQALAEKDLIVERANSR--VEVVHERL-----SSLENO 910
   || || || || || || || || || || || || || || || || || || ||
76 ATHPDVMDVFPVFPQKREIFGELVTKHOLRTSFHPNQFTLTSPKESVTKNVDTMAYH 135
   || || || || || || || || || || || || || || || || || || ||
911 VTIADEKYEFL--YAEKOSIEEDLANKQTEISYLSDLSTLEKLSIKKDKQT--ISSK 966
   || || || || || || || || || || || || || || || || || || ||
136 YRMLEAMGTADRSVINI-----HIGGAYGNKDTATQFQHNKOLPOET----- 179
   || || || || || || || || || || || || || || || || || || ||
967 YKELE-----KDYLNIMADYQHSQHLNLEKAIINEKNINIRELNEKMLRLDDELLKKQ 1020
   || || || || || || || || || || || || || || || || || || ||
180 -----KERMTLENDKTYTT-----BETLQVCEQEDVPFV 209
   || || || || || || || || || || || || || || || || || || ||
1021 RSYDTKVOQLREENASLKDCQRTYESOLASLSVSKYSESELNKKEALVIFQKEITEYR 1080
   || || || || || || || || || || || || || || || || || || ||
210 FDFHHFYANPDHDLNVALPRMIKWTWERICLOPKVHLSPPKSEQAIRS--HADVVDANFL 268
   || || || || || || || || || || || || || || || || || || ||
1081 DQLHKAFQNPETHININDVKSGLNSDENIYSTSTLTSLKDVQLKSLHTK--EANQL 1138
   || || || || || || || || || || || || || || || || || || ||
269 LERFRQWGTNIDFMIEAKOKDKALLRLMDELSSI 302
   || || || || || || || || || || || || || || || || || || ||
1139 SERIKEISEMLEQSIATEEK---LRRKNSLDCI 1169

RESULT 13
AP04_MACFA STANDARD; PRT: 429 AA.
AC P33621;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE APOLIPROTEIN A-IV PRECURSOR (APO-AIV).
GN APOA4.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Leukocyte;
RX MEDLINE=93192330; PubMed=8448212;
RA Osada J., Pocovi M., Nicolosi R.J., Schaefer E.J., Ordovas J.M.;
RT "Nucleotide sequences of the Macaca fascicularis apolipoprotein C-III
RT and A-IV genes.";
RL Biochim. Biophys. Acta 1172:335-339(1993).
CC -!- FUNCTION: MAY HAVE A ROLE IN CHYLOMICRONS AND VLDL SECRETION AND
CC CATABOLISM. REQUIRED FOR EFFICIENT ACTIVATION OF LIPOPROTEIN
CC LIPASE BY APOC-II; POTENT ACTIVATOR OF LCAT. APOA-IV IS A MAJOR
CC COMPONENT OF HDL AND CHYLOMICRONS.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -!- TISSUE SPECIFICITY: SECRETED IN PLASMA.
CC -!- DOMAIN: NINE OF THE THIRTEEN 22-AMINO ACID TANDEM REPEATS (EACH
CC 22-MER IS ACTUALLY A TANDEM ARRAY OF TWO, A AND B, RELATED 11-
CC MERS) OCCURRING IN THIS SEQUENCE ARE PREDICTED TO BE HIGHLY ALPHA-
CC HELICAL, AND MANY OF THESE HELICES ARE AMPHIPATHIC. THEY MAY
CC THEREFORE SERVE AS LIPID-BINDING DOMAINS WITH LECITHIN:CHOLESTEROL
CC ACYLTRANSFERASE (LCAT) ACTIVATING ABILITIES.
CC -!- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X68361; CAA48421.1; -
CC PIR; S29565; S29565.
CC PIR; S30195; S30195.
CC HSSP; P02649; INFO.

```

Query Match 5.4%; Score 90; DB 1; Length 2104;
Best Local Similarity 19.8%; Pred. No. 40;
Matches 66; Conservative 58; Mismatches 130; Indels 80; Gaps 12;

DR	InterPro: IPR000074; Apolipoprotein.	
DR	Pham; PF01442; Apolipoprotein; 1.	
KW	Plasma; Lipid transport; HDL; Chylomicron; Repeat; Signal.	
FT	SIGNAL	1 20 BY SIMILARITY.
FT	CHAIN	21 429 APOLIPOPROTEIN A-IV.
FT	DOMAIN	33 330 13 X 22 AA APPROXIMATE TANDEM REPEATS.
FT	REPEAT	33 54 1.
FT	REPEAT	60 81 2.
FT	REPEAT	82 103 3.
FT	REPEAT	115 136 4.
FT	REPEAT	137 158 5.
FT	REPEAT	159 180 6.
FT	REPEAT	181 202 7.
FT	REPEAT	203 224 8.
FT	REPEAT	225 246 9.
FT	REPEAT	247 268 10.
FT	REPEAT	269 286 11.
FT	REPEAT	287 308 12.
FT	REPEAT	309 330 13.
FT	DOMAIN	372 420 GLU/GLN-RICH.
SQ	SEQUENCE	429 AA; 49876 MW; 3D458F51D0DB60C CRC64;

Query Match	5.4%;	Score 89.5;	DB 1;	Length 429;
Best Local Similarity	20.4%;	Pred. No. 5.2;	Indels	65; Gaps
Matches	68;	Conservative	59;	Mismatches 142;

QY	8	VSNAMSLWDASPAKTLTTFARYSKLSKTERKEALLTVTKANLRNTMRTLHYIIGHGIPLYR	67
Db	15	VTGAREVSDAVQATVMWDFYSQLS-SNAKEAVEHLQKSELQTQNLNLFQDKLGEVNTYA	73
QY	68	--FSSSIIVPLAT--HPDVMWD-----FVTPFOKEPR-----IGELVKTHQL	105
Db	74	GDLOKKLPFATELHERLAKDSEKLEETIRKELEEVRABLPHANEVSQKTIGENVRELQ	133
QY	106	RTSFHPNQPTLFTSPKESVTKNAVTDMAYHY-RLMEAMGIADRSVINIHIGAYGNKDTA	164
Db	134	RLPEVTDQLRTQVNTQTLELRRLQTPYAORMERVLRNADSLQTSLRPH-----ADQL	186
QY	165	TAQFHNKIQLPOLKEKRWMLTENDDKTYTTEETLQVCEQEDVPFVDFPHFYANPDHAD	224
Db	187	KAKIDQNV-----BELKRLTPYADEFKVIDQTVLELRSLAPYA-----QDAQEK	233
QY	225	LNVALP----RMKWTWRIGLPQVHLSSPKSFOAIRSHADYVDANF-----LLERFRQ	274
Db	234	LNHOLEGLAFQMKNAEE--LKARISAAEELRQRIAPLAEDMRGNLRNTEGLQKSLAE	291
QY	275	WGTNTIDFIE-----AKQDKALLRLMDEL	299
Db	292	LGCHLDRHVEEFLRVEPYGENFNKALVQOMEOL	325

RESULT	14	
ID	BACA_BACLI	
AC	STANDARD;	PRT; 5255 AA.
OC	O68006;	
DC	15-JUL-1999	(Rel. 38, Created)
DT	15-JUL-1999	(Rel. 38, Last sequence update)
DT	20-AUG-2001	(Rel. 40, Last annotation update)
DE	BACTRACIN SYNTHETASE 1 (BAC) [INCLUDES: ATP-DEPENDENT ISOLEUCINE ADENYLASE (ILEA) (ISOLEUCINE ACTIVASE); ATP-DEPENDENT CYSTEINE ADENYLASE (CYSA) (CYSTEINE ACTIVASE); ATP-DEPENDENT LEUCINE ADENYLASE (LEUA) (LEUCINE ACTIVASE); ATP-DEPENDENT GLUTAMATE ADENYLASE (GLUA) (GLUTAMATE ACTIVASE); ATP-DEPENDENT ISOLEUCINE ADENYLASE (ILEA) (ISOLEUCINE ACTIVASE); GLUTAMATE RACEMASE (EC 5.1.1.3)].	
GN	BACA.	
OS	Bacillus licheniformis.	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group.	
OC	Bacillus/Staphylococcus group; Bacillus.	
NCBI_taxid=1402;		
[[1]		
RN		
RP	SEQUENCE FROM N.A.	
RP	STRAIN=ATCC 10716.	

MEDLINE=98089193; PubMed=9427658;
Konz D., Kiens A., Schoergendorfer K., Marahiel M.A.;
"The bacitracin biosynthesis operon of Bacillus licheniformis ATCC
10716: molecular characterization of three multi-modular peptide
synthetases.";
Chem. Biol. 4:927-937(1997).
CC -!- FUNCTION: ACTIVATES FIVE AMINO ACIDS, INCORPORATES TWO D-AMINO
CC ACIDS, RELEASES AND CYCLIZES THE MATURE BACITRACIN.
CC -!- CATALYTIC ACTIVITY: L-GLUTAMATE = D-GLUTAMINE.
CC -!- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTHETHEINES
CC (POTENTIAL).
CC -!- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF THE CYCLIC PEPTIDE
CC ANTIBIOTIC BACITRACIN.
CC -!- SUBUNIT: LARGE MULTIEZYME COMPLEX OF BA1, BA2 AND BA3.
CC -!- DOMAIN: CONSISTS OF FIVE MODULES AND ONE EPIMERIZATION DOMAIN IN
CC THE FOURTH MODULE. EACH MODULE INCORPORATES ONE AMINO ACID INTO
CC THE PEPTIDE PRODUCT AND CAN BE FURTHER SUBDIVIDED INTO DOMAINS
CC RESPONSIBLE FOR SUBSTRATE ADENYLATION, THIOLATION, CONDENSATION
CC (NOT FOR THE INITIATION MODULE), AND EPIMERIZATION (OPTIONAL), AND
CC N METHYLATION (OPTIONAL).
CC -!- MISCELLANEOUS: BACITRACIN IS A MIXTURE OF AT LEAST TEN CYCLIC
CC DODECAPEPTIDES, THAT DIFFER BY ONE OR TWO AMINO ACIDS. THE MOST
CC ABUNDANT IS BACITRACIN A, A BRANCHED CYCLIC DODECAPEPTIDE. IT
CC CONTAINS AN N-TERMINAL LINEAR PENTAPEPTIDE MOIETY (ILE-CYS-LEU-D-
CC GLU-ILE) WITH AN ISOLEUCINE-CYSTEINE THIAZOLINE CONDENSATION
CC PRODUCT AND A C-TERMINAL HEPTAPEPTIDE RING (LYS-D-ORN-ILE-D-PHE-
CC HIS-D-ASP-ASN), IN WHICH THE FREE ALPHA-CARBOXY GROUP OF THE C-
CC TERMINAL ASN IS BOUND TO THE EPSILON-AMINO GROUP OF LYSINE. IT
CC CONTAINS FOUR AMINO ACIDS IN THE D-CONFIGURATION (GLU-4, ORN-7,
CC PHE-9, AND ASP-11).
CC -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
or send an email to license@isb-sib.ch).

EMBL: AF007865; AAC06346.1; -;
DR InterPro: IPR000873; AMP-bind.
DR InterPro: IPR001242; DUF4.
DR InterPro: IPR003880; Phosphopant_attach.
DR Pfam: PF00501; AMP-binding; 5.
DR Pfam: PF00668; Condensation; 5.
DR Pfam: PF00550; pp-binding; 5.
DR PRINTS: PR00154; AMPBINDING.
DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; 3.
DR PROSITE: PS00455; AMP-BINDING; 5.
DR PROSITE: PS50075; ACP_DOMAIN; 5.
KW Ligase; Isomerase; Antibiotic biosynthesis; Phosphopantetheine;
KW Multifunctional enzyme; Repeat.
KW REPEAT 39 612
FT REPEAT 1109 1648
FT REPEAT 2124 2689
FT REPEAT 3164 3732
FT REPEAT 4668 5249
FT DOMAIN 621 1037
FT DOMAIN 544 612
FT DOMAIN 1585 1648
FT DOMAIN 2621 2689
FT DOMAIN 3664 3732
FT DOMAIN 5171 5249
FT BINDING 574 574
FT BINDING 1615 1615
FT BINDING 2651 2651
FT BINDING 3694 3694
FT BINDING 5201 5201
FT SEQUENCE 5225 AA; 598254 MW; 906E8D068450F85B CRC64.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:51:31 ; Search time 45.07 Seconds
(without alignments)
673.585 Million cell updates/sec

Title: US-09-724-296-2

Perfect score: 4349

Sequence: 1 MTKPLTIGYWKIKGLVQPTR.....KELTARKRSRKEEVEDEK 828

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1170	26.9	218	1	GT26_SCHJA
2	984	22.6	218	1	GT27_SCHMA
3	921	21.2	218	1	GT26_SCHMA
4	710.5	16.3	220	1	GT29_FASHE
5	699	16.1	217	1	GT27_FASHE
6	681	15.7	217	1	GT28_FASHE
7	680	15.6	217	1	GT26_FASHE
8	481.5	11.1	217	1	GT26_FASHE
9	481.5	11.1	217	1	GT26_FASHE
10	477.5	11.0	217	1	GT26_FASHE
11	474.5	10.9	218	1	GT26_FASHE
12	473.5	10.9	217	1	GT26_FASHE
13	473.5	10.9	217	1	GT26_FASHE
14	468.5	10.8	217	1	GT26_FASHE
15	468.5	10.8	217	1	GT26_FASHE
16	467.5	10.7	217	1	GT26_FASHE
17	465.5	10.7	217	1	GT26_FASHE
18	463.5	10.7	217	1	GT26_FASHE
19	461.5	10.6	217	1	GT26_FASHE
20	455.5	10.5	217	1	GT26_FASHE
21	452.5	10.4	217	1	GT26_FASHE
22	447.5	10.3	225	1	GT26_FASHE
23	437.5	10.1	224	1	GT26_FASHE
24	432.5	9.9	219	1	GT26_FASHE
25	385.5	8.9	219	1	GT26_FASHE
26	360	8.3	320	1	GT26_FASHE
27	248.5	5.7	208	1	GT26_FASHE
28	245.5	5.6	208	1	GT26_FASHE
29	244	5.6	209	1	GT26_FASHE
30	242	5.6	207	1	GT26_FASHE
31	240	5.5	209	1	GT26_FASHE
32	238	5.5	209	1	GT26_FASHE
33	237	5.4	209	1	GT26_FASHE

34 237 5.4 209 1 GTP_RAT
35 236 5.4 209 1 GTP_CRILLO
36 235 5.4 209 1 GTP_MACMU
37 233 5.4 209 1 GTP_MOUSE
38 230 5.3 209 1 GTP_BOVIN
39 222.5 5.1 223 1 GTP_RABIT
40 219.5 5.0 208 1 GTP_CARTEL
41 200.5 4.6 221 1 GTP_RAT
42 198.5 4.6 221 1 GTP_MOUSE
43 198.5 4.6 222 1 GTP_MOUSE
44 194.5 4.5 221 1 GTP_RAT
45 193.5 4.4 220 1 GTP_RAT

ALIGNMENTS

RESULT 1
GT26_SCHJA
ID GT26_SCHJA STANDARD; PRT; 218 AA.
AC P08515;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE GLUTATHIONE S-TRANSFERASE 26 KDA (EC 2.5.1.18) (GST 26) (SJ26 ANTIGEN)
DE (GST CLASS-ALPHA).
OS Schistosoma japonicum (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Turbellarian Platyhelminths;
OC Rhabdiphora; Eulicthophora; Reverteospermatia; Mediofusata;
OC Neodermata; Trematoda; Digenea; Strigeidida; Schistosomatoidea;
OC Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6182;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=870411520; PubMed=3095841;
RA Smith D.B., Davern K.M., Board P.G., Tiu W.U., Garcia E.G.,
RA Mitchell G.F.;
RT "Mr 26,000 antigen of Schistosoma japonicum recognized by resistant
WEHI 129/J mice is a parasite glutathione S-transferase.";
Proc. Natl. Acad. Sci. U.S.A. 83:8703-8707(1986).
[2]
REVIEWS.
RA Smith D.B., Davern K.M., Board P.G., Tiu W.U., Garcia E.G.,
RA Mitchell G.F.;
Proc. Natl. Acad. Sci. U.S.A. 84:6541-6541(1987).
[3]
X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=95276631; PubMed=7538846;
Lim K., Ho J.X., Keeling K., Gilliland G.L., Ji X., Ruker F.,
Carter D.C.;
"Three-dimensional structure of Schistosoma japonicum glutathione S-
transferase fused with a six-amino acid conserved neutralizing
epitope of gp41 from HIV.";
Protein Sci. 3:2233-2244(1994).
[4]
X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RX MEDLINE=95156484; PubMed=7853399;
McTigue M.A., Williams D.R., Tainer J.A.;
"Crystalline structures of a schistosomal drug and vaccine target:
glutathione S-transferase from Schistosoma japonica and its complex
with the leading antischistosomal drug praziquantel.";
J. Mol. Biol. 246:21-27(1995).
CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -!- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
HAEMATIN IN THE PARASITE GUT.
CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE -> HX + R-S-GUTATHIONE.
CC -!- SUBUNIT: HOMODIMER.
CC -!- MISCELLANEOUS: THERE ARE AT LEAST TWO ISOENZYMES OF GST IN
S. JAPONICUM.
CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M14654; AAB59203.1; -;
 DR PIR; A26484; A26484.
 DR PDB; 1GNE; 30-NOV-94.
 DR PDB; 1GTA; 07-FEB-95.
 DR PDB; 1GTH; 01-DEC-95.
 DR PDB; 1B8X; 12-APR-99.
 DR InterPro; IPR000521; GST.
 DR Pfam; PF00043; GST; 1.
 DR Transferase; Antigen; Multigene family; 3D-structure.
 KW SEQUENCE 218 AA; 25498 MW; 5E2AC418BD0EF13F CRC64;

Query Match 26.9%; Score 1170; DB 1; Length 218;
 Best Local Similarity 100.0%; Pred. No. 2.8e-56;
 Matches 216; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 PILGYWKIKGLVQPTRLLEYLEEKYEHLRYERDEGDKWRNKKFELGLEFPNLPYYIDGD 64
 DB 3 PILGYWKIKGLVQPTRLLEYLEEKYEHLRYERDEGDKWRNKKFELGLEFPNLPYYIDGD 62
 QY 65 VKLTQSMARIYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRISYKDFETLKVD 124
 DB 63 VKLTQSMARIYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRISYKDFETLKVD 122
 QY 125 LSKLPEMLKMFEDRLCHKTYLNGDHTVDFDMLYDLDVLYMDPCLDAPFKLVCFKRR 184
 DB 123 LSKLPEMLKMFEDRLCHKTYLNGDHTVDFDMLYDLDVLYMDPCLDAPFKLVCFKRR 182
 QY 185 IEAIPQIDKYLKSSKYIAWPLQGWQATFGGDDHPPK 220
 DB 183 IEAIPQIDKYLKSSKYIAWPLQGWQATFGGDDHPPK 218

RESULT 2
 GT27_SCHMA STANDARD; PRT; 218 AA.
 ID GT27_SCHMA STANDARD; PRT; 218 AA.
 AC P35661;
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DE GLUTATHIONE S-TRANSFERASE 26 KDA (EC 2.5.1.18) (GST 26) (SM26/2)
 DE ANTIGEN (GST CLASS-ALPHA).
 OS Schistosoma mansoni (Blood fluke).
 OC Eukaryota; Metazoa; Platyhelminthes; Turbellarian Platyhelminths;
 OC Rhabditophora; Eulicthophora; Revertospermata; Mediofusata;
 OC Neodermata; Trematoda; Digenea; Strigeidida; Schistosomatoidea;
 OC Schistosomatidae; Schistosoma.
 OX NCBI_TaxID=6183;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92131046; PubMed=1775156;
 RA Wright M.D., Harrison R.A., Melider A.M., Newport G.R., Mitchell G.F.;
 RT "Another 26-kilodalton glutathione S-transferase of Schistosoma
 RT mansoni".
 RL Mol. Biochem. Parasitol. 49:177-179(1991).
 CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -!- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
 CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
 CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
 CC HAEMATIN IN THE PARASITE GUT.
 CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE = HX + R-S-GLUTATHIONE.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M73624; -; NOT_ANNOTATED_CDS.
 DR PIR; A45556; A45556.
 DR HSP; P08515; IGTB.
 DR InterPro; IPR000521; GST.
 DR Pfam; PF00043; GST; 1.
 DR Transferase; Antigen; Multigene family.
 KW SEQUENCE 218 AA; 25411 MW; D803EE9028B36185 CRC64;

Query Match 22.6%; Score 984; DB 1; Length 218;
 Best Local Similarity 82.9%; Pred. No. 2.5e-46;
 Matches 179; Conservative 18; Mismatches 19; Indels 0; Gaps 0;
 QY 5 PILGYWKIKGLVQPTRLLEYLEEKYEHLRYERDEGDKWRNKKFELGLEFPNLPYYIDGD 64
 DB 3 PKLYGWKIKGLVQPTRLLEYLEEKYEHLRYERDEGDKWRNKKFELGLEFPNLPYYIDGD 62
 QY 65 VKLTQSMARIYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRISYKDFETLKVD 124
 DB 63 VKLTQSMARIYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRISYKDFETLKVD 122
 QY 125 LSKLPEMLKMFEDRLCHKTYLNGDHTVDFDMLYDLDVLYMDPCLDAPFKLVCFKRR 184
 DB 123 LSKLPEMLKMFEDRLCHKTYLNGDHTVDFDMLYDLDVLYMDPCLDAPFKLVCFKRR 182
 QY 185 IEAIPQIDKYLKSSKYIAWPLQGWQATFGGDDHPPK 220
 DB 183 IEAIPQIDKYLKSSKYIAWPLQGWQATFGGDDHPPK 218

RESULT 3
 GT26_SCHMA STANDARD; PRT; 218 AA.
 ID GT26_SCHMA STANDARD; PRT; 218 AA.
 AC P15964;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE GLUTATHIONE S-TRANSFERASE 26 KDA (EC 2.5.1.18) (GST 26) (SM26/1)
 DE ANTIGEN (GST CLASS-ALPHA).
 OS Schistosoma mansoni (Blood fluke).
 OC Eukaryota; Metazoa; Platyhelminthes; Turbellarian Platyhelminths;
 OC Rhabditophora; Eulicthophora; Revertospermata; Mediofusata;
 OC Neodermata; Trematoda; Digenea; Strigeidida; Schistosomatoidea;
 OC Schistosomatidae; Schistosoma.
 OX NCBI_TaxID=6183;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=PUERTO RICAN;
 RX MEDLINE=90348716; PubMed=2385266;
 RA Trottein F., Kieny M.P., Verwaerde C., Torpier G., Pierce R.J.,
 RA Balloul J.-M., Schmitt D., Lecocq J.-P., Capron A.;
 RT "Molecular cloning and tissue distribution of a 26-kilodalton
 RT glutathione S-transferase of Schistosoma japonicum and Schistosoma
 RL mansoni".
 RL Mol. Biochem. Parasitol. 41:35-44(1990).
 RN [2]
 RP SEQUENCE OF 8-218 FROM N.A.
 RC STRAIN=PUERTO RICAN;
 RX MEDLINE=90271935; PubMed=1693415;
 RA Henkle K.J., Davern K.M., Wright M.D., Ramos A.J., Mitchell G.F.;
 RT "Comparison of the cloned genes of the 26- and 28-kilodalton
 RT glutathione S-transferases of Schistosoma japonicum and Schistosoma
 RT mansoni".
 RL Mol. Biochem. Parasitol. 40:23-34(1990).
 CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER

CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -!- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
 CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
 CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
 CC HAEMATIN IN THE PARASITE GUT.
 CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE = HX + R-S-GLUTATHIONE.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- TISSUE SPECIFICITY: TEGUMENT AND IN SUBTEGUMENTARY PARENCHYMAL
 CC CELLS. GST 26 MAY BE ACTIVELY EXCRETED BY ADULT WORMS.
 CC -!- MISCELLANEOUS: THERE ARE AT LEAST THREE ISOENZYMES OF GST IN
 CC S.MANSONI.
 CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

CC EMBL: M31106; AAA29888.1; -;
 CC EMBL: M26913; AAA29889.1; -;
 CC PIR: A45523; A45523.
 CC HSP: P08515; IGTB.
 CC InterPro: IPR000521; GST.
 CC Pfam: PF00043; GST; 1.
 CC Transferase: Antigen: Multigene family.
 CC SEQUENCE 218 AA; 25401 MW; 061A6548A842D6E8 CRC64;

Query Match 21.2%; Score 921; DB 1; Length 218;
 Best Local Similarity 80.1%; Pred. No. 5.8e-43;
 Matches 173; Conservative 14; Mismatches 29; Indels 0; Gaps 0;
 QY 5 PILGYWIKGLVQPTRLLEYLEEKYEHELYERDEGDKRNKKEFELGEPNLPYYIDGD 64
 Db 3 PKGYWYKGLVQPTRLLEHLEETEERAYDRNEIDAWSNDKFKLGLPEPNLPYYIDGD 62
 QY 65 VKLTQSMAIRIYADKHNMLGCGPKRAEISMLEGAVLDIRYGVSRITAYSKDFETLKVDF 124
 Db 63 FKLTQSMAIRIYADKHNMLGCGPKRAEISMLEGAVLDIRMGVLRITAYKNEYETLKVDF 122
 QY 125 LSKPEMLKMFEDRLCHTKTYLNGDHVTHPDPMYLDALDVLVYMDPCLDAFPKLVCFKKR 184
 Db 123 LNKLPGLRKMFEDRLSNKTYLNGCVTHPDPMYLDALDVLVYMDPCLDAFPKLVCFKK 182
 QY 185 IEATPQIDKYLKSSKYTAWPLQGWQATFGGDDHPPK 220
 Db 183 IEDLPQIKNVLNSSRIKWPLOGWDATFGGDDTPPK 218

RESULT 4
 GT29_FASHE STANDARD; PRT; 220 AA.
 AC P56598;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE GLUTATHIONE S-TRANSFERASE 26 KDA 1 (EC 2.5.1.18) (GST1) (FH1) (GST
 DE CLASS-ALPHA).
 OS Fasciola hepatica (Liver fluke).
 OC Eukaryota; Metazoa; Platyhelminthes; Turbellarian Platyhelminths;
 OC Rhabditophora; Eulcithophora; Revertospermata; Mediofusata;
 OC Neodermata; Trematoda; Digenea; Echinostomida; Echinostomata;
 OC Fasciolidae; Fasciolidae; Fasciola.
 OX NCBI_TaxID=6192;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=92155306; PubMed=1740183;
 RA Panaccio M., Wilson L.R., Cramer S.L., Wijffels G.L., Spithill T.W.;
 RT "Molecular characterization of cDNA sequences encoding glutathione S-
 RT transferases of Fasciola hepatica."

RL Exp. Parasitol. 74:232-237(1992).
 RN [2]
 RP ERRATUM.
 RX MEDLINE=94039664; PubMed=8224094;
 RA Panaccio M., Wilson L.R., Cramer S.L., Wijffels G.L., Spithill T.W.;
 RL Exp. Parasitol. 77:385-385(1993).
 RN [3]
 RP SEQUENCE OF 22-220 FROM N.A.
 RA Cramer S.;
 RL Patent number WO9008819, 09-AUG-1990.
 CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -!- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
 CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
 CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
 CC HAEMATIN IN THE PARASITE GUT.
 CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE = HX + R-S-GLUTATHIONE.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

CC EMBL: A00993; CAA00118.1; -;
 CC InterPro: IPR000521; GST.
 CC Pfam: PF00043; Antigen: Multigene family.
 CC Transferase: Antigen: Multigene family.
 CC INIT_MET 0 BY SIMILARITY.
 CC CONFLICT 22 22 Y -> V (IN REF. 3).
 CC CONFLICT 110 111 DP -> VS (IN REF. 3).
 CC CONFLICT 189 189 A -> P (IN REF. 3).
 CC SEQUENCE 220 AA; 25598 MW; 27B9F150B75D101F CRC64;

Query Match 16.3%; Score 710.5; DB 1; Length 220;
 Best Local Similarity 59.7%; Pred. No. 1.1e-51;
 Matches 129; Conservative 32; Mismatches 54; Indels 1; Gaps 1;
 QY 7 LGWKIKGLVQPTRLLEYLEEKYEHELYERDEGDKRNKKEFELGEPNLPYYIDGDVK 66
 Db 4 LGWKIRGLVQPTRLLEYLEEKYEHELYERDEGDKRNKKEFELGEPNLPYYIDDKCK 62
 QY 67 LTQSMAIRIYADKHNMLGCGPKRAEISMLEGAVLDIRYGVSRITAYSKDFETLKVDFLS 126
 Db 63 LTQSLAIRIYADKHNMLGCGPKRAEISMLEGAVLDIRYGVSRITAYSKDFETLKVDFLS 122
 QY 127 KLPEMLKMFEDRLCHTKTYLNGDHVTHPDPMYLDALDVLVYMDPCLDAFPKLVCFKKRIE 186
 Db 123 DLPTTKMWSDFLGKNPRLGTSVSHVDPVMEALDAIRYLEPHCLDHPNLPQQFMSRIE 182
 QY 187 AIPOIDKYLKSSKYTAWPLQGWQATFGGDDHPPKSD 222
 Db 183 ALPSKAYMESNRNFKWPLNGWHAQFGGDDAPPSE 218

RESULT 5
 GT27_FASHE STANDARD; PRT; 217 AA.
 ID GT27_FASHE
 AC P31670;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE GLUTATHIONE S-TRANSFERASE 26 KDA 47 (EC 2.5.1.18) (GST47) (FH47)
 DE (GST CLASS-ALPHA).
 OS Fasciola hepatica (Liver fluke).
 OC Eukaryota; Metazoa; Platyhelminthes; Turbellarian Platyhelminths;
 OC Rhabditophora; Eulcithophora; Revertospermata; Mediofusata;
 OC Neodermata; Trematoda; Digenea; Echinostomida; Echinostomata;

```

OC Fasciolidae; Fasciolidae; Fasciola.
OX NCBI_TaxID=6192;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92155306; PubMed=1740183;
RA Panaccio M., Wilson L.R., Crameri S.L., Wijffels G.L., Spithill T.W.;
RT "Molecular characterization of cDNA sequences encoding glutathione S-
transferases of Fasciola hepatica."
RL Exp. Parasitol. 74:232-237(1992).
RN [2]
RP ERRATUM.
RX MEDLINE=94039664; PubMed=8224094;
RA Panaccio M., Wilson L.R., Crameri S.L., Wijffels G.L., Spithill T.W.;
RL Exp. Parasitol. 77:385-385(1993).
RN [3]
RP SEQUENCE OF 7-105 FROM N.A.
RA Crameri S.;
RL Patent number WO9008819, 09-AUG-1990.
RN [4]
RP X-RAY CRYSTALLOGRAPHY (3 ANGSTROMS).
RX MEDLINE=98035725; PubMed=9367777;
RA Rossjohn J., Feil S.C., Wilce M.C.J., Sexton J.L., Spithill T.W.,
RT "Crystallization, structural determination and analysis of a novel
parasitic vaccine candidate: Fasciola hepatica glutathione
S-transferase."
RL J. Mol. Biol. 273:857-872(1997).
CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -!- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
HAEMATIN IN THE PARASITE GUT.
CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE = HX + R-S-GLUTATHIONE.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL: M77681; AAA29140.1; -.
CC EMBL: A00996; CAA00121.1; -.
CC PDB: 1PHE; 29-JUL-98.
CC InterPro; IPR000521; GST.
CC Pfam; PF00043; GST; 1.
CC Transferase; Antigen; Multigene family; 3D-structure.
CC INIT_MET 0
CC CONFLICT 65 65 T -> I (IN REF. 3).
CC CONFLICT 102 105 RIGF -> FEEL (IN REF. 3).
CC SEQUENCE 217 AA; 25281 MW; 0FB8BBFE63029E03 CRC64;

Query Match 16.1%; Score 699; DB 1; Length 217;
Best Local Similarity 58.2%; Pred. No. 4.3e-31;
Matches 124; Conservative 36; Mismatches 53; Indels 0; Gaps 0;

QY 7 LGYWKIKGLVQPTKLLLEYLEEKYEHEHYERDEGDKWRNKKFELGFPNLPYYIDGDKV 66
DB 4 LGYWKURGLAQPVRLFLYLEYGLGEHYHLYGRDDRDKWSEKFNMGDLPLNLPYYIDDKCK 63

QY 67 LTQSMATIRYIADKHNKMLGCGPKEAETSMLEGAVALDTRYGVSRISYKDFETLKVDFLS 126
DB 64 LTQSVAIMRYIADKHKMLGTTTPEERARISMIEGAANDLRIGRVCYNPKFEVKEEVK 123

QY 127 KLPEMLKMFEDRLCHTKYUNGDHVTHPDPMYDALDVLVYMPDMLCDPAPKLVCFKKK 186
DB 124 ELPEKTLKMSDFLGRHYLTGGSVSHVDFMFLYETLDSYRLAPHCLDFEPKLFKFSR 183

```

```

QY 187 AIPDIDKYLKSSKYIAMPLOQWOATEGGDHPP 219
DB 184 ALPKIKAYMESKRFJKWPLNGWAASFAGDAPP 216

```

```

RESULT 6
GT28_FASHE STANDARD; PRT: 217 AA.
AC P31671;
DT 01-JUL-1993 (Rel. 26, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GLUTATHIONE S-TRANSFERASE 26 KDA 7 (EC 2.5.1.18) (GST7) (FH7) (GST
DE CLASS-ALPHA).
OS Fasciola hepatica (Liver fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Turbellarian Platyhelminths;
OC Rhabditophora; Eulicthophora; Revertospermata; Mediofusata;
OC Neodermata; Trematoda; Digenea; Echinostomida; Echinostomata;
OC Fasciolidae; Fasciolidae; Fasciola.
OX NCBI_TaxID=6192;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92155306; PubMed=1740183;
RA Panaccio M., Wilson L.R., Crameri S.L., Wijffels G.L., Spithill T.W.;
RT "Molecular characterization of cDNA sequences encoding glutathione S-
transferases of Fasciola hepatica."
RL Exp. Parasitol. 74:232-237(1992).
RN [2]
RP ERRATUM.
RX MEDLINE=94039664; PubMed=8224094;
RA Panaccio M., Wilson L.R., Crameri S.L., Wijffels G.L., Spithill T.W.;
RL Exp. Parasitol. 77:385-385(1993).
RN [3]
RP SEQUENCE OF 8-217 FROM N.A.
RA Crameri S.;
RL Patent number WO9008819, 09-AUG-1990.
CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -!- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
HAEMATIN IN THE PARASITE GUT.
CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE = HX + R-S-GLUTATHIONE.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL: M77680; AAA29139.1; -.
CC EMBL: A00994; CAA00119.1; -.
CC HSP; P08515; IGTB.
CC InterPro; IPR000521; GST.
CC Pfam; PF00043; GST; 1.
CC Transferase; Antigen; Multigene family.
CC INIT_MET 0
CC CONFLICT 35 42 NOREKWL -> MIGENGA (IN REF. 3).
CC CONFLICT 188 196 IKYMKSR -> SRYMSRA (IN REF. 3).
CC SEQUENCE 217 AA; 25196 MW; 0099E1F59E49A49E CRC64;

```

```

Query Match 15.7%; Score 681; DB 1; Length 217;
Best Local Similarity 58.7%; Pred. No. 4e-30;
Matches 125; Conservative 29; Mismatches 59; Indels 0; Gaps 0;

QY 7 LGYWKIKGLVQPTKLLLEYLEEKYEHEHYERDEGDKWRNKKFELGFPNLPYYIDGDKV 66
DB 124 ELPEKTLKMSDFLGRHYLTGGSVSHVDFMFLYETLDSYRLAPHCLDFEPKLFKFSR 183

```


Acta Crystallogr. D 50:219-224(1994).
 [9]
 RN X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=9415386; PubMed=8110735;
 JI X., Johnson W.W., Sesay M.A., Dickert L., Prasad S.M., Ammon H.L.,
 RA Armstrong R.N., Gilliland G.L.;
 RA "Structure and function of the xenobiotic substrate binding site of a
 RT glutathione S-transferase as revealed by X-ray crystallographic
 RT analysis of product complexes with the diastereomers of 9-(S-
 RT glutathionyl)-10-hydroxy-9,10-dihydrophenanthrene.";
 RL Biochemistry 33:1043-1052(1994).
 CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -!- FUNCTION: THE OLFACTORY GST MAY BE CRUCIAL FOR THE ACUITY OF THE
 CC OLFACTORY PROCESS.
 CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE = HX + R-S-GLUTATHIONE.
 CC -!- SUBUNIT: HOMODIMER OR HETERODIMER.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- MISCELLANEOUS: RAT LIVER GLUTATHIONE S-TRANSFERASES (GST) CONSIST
 CC OF BINARY COMBINATIONS OF THREE MAJOR CLASSES OF SUBUNITS
 CC DESIGNATED: YA (25 KDA), YB (27 KDA) AND YC (28 KDA).
 CC -!- MISCELLANEOUS: YB SUBCLASS SELECTIVELY BINDS STEROID HORMONES.
 CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X04229; CAA27811.1; -;
 DR EMBL; J12719; AAA41287.1; -;
 DR EMBL; J02810; AAA41293.1; -;
 DR PIR; A24085; A24085.
 DR PIR; A25510; A25510.
 DR PIR; A29794; A29794.
 DR PIR; S17167; S17167.
 DR PDB; 1GSB; 31-OCT-93.
 DR PDB; 1GSC; 31-OCT-93.
 DR PDB; 2GST; 31-OCT-93.
 DR PDB; 3GST; 31-JAN-94.
 DR PDB; 4GST; 31-OCT-93.
 DR PDB; 5GST; 31-OCT-93.
 DR PDB; 6GST; 08-NOV-96.
 DR PDB; 6GSU; 08-NOV-96.
 DR PDB; 6GSV; 08-NOV-96.
 DR PDB; 6GSW; 08-NOV-96.
 DR PDB; 6GSX; 08-NOV-96.
 DR PDB; 6GSY; 08-NOV-96.
 DR PDB; 5FWG; 27-JAN-99.
 DR InterPro; IPR000521; GST.
 DR InterPro; IPR003081; GST_mu.
 DR Pfam; PF00043; GST; 1.
 DR PRINTS; PR01267; GSTRSFRASEM.
 DR Transferase; Multigene family; 3D-structure; Olfaction.
 KW INIT_MET 0
 FT MUTAGEN 86 C->S: NO CHANGE IN ACTIVITY.
 FT CONFLICT 168 I -> N (IN REF. 3).
 FT CONFLICT 198 KS -> NC (IN REF. 2).
 FT STRAND 2 7
 FT TURN 11 13
 FT HELIX 14 22
 FT TURN 23 24
 FT STRAND 27 32
 FT TURN 37 39
 FT HELIX 43 46
 FT TURN 47 50
 FT STRAND 61 64
 FT TURN 65 66
 FT STRAND 67 70
 FT HELIX 72 82

FT TURN 83 84
 FT HELIX 90 114
 FT TURN 115 118
 FT TURN 117 118
 FT HELIX 119 128
 FT TURN 129 129
 FT HELIX 130 141
 FT TURN 142 143
 FT TURN 154 154
 FT HELIX 155 169
 FT TURN 171 176
 FT HELIX 178 188
 FT TURN 189 189
 FT HELIX 191 196
 FT TURN 197 198
 FT TURN 200 201
 FT TURN 210 211
 SQ SEQUENCE 217 AA; 25782 MW; 2ACE8D49DA785118 CRC64;
 Query Match 11.1%; Score 481.5; DB 1; Length 217;
 Best Local Similarity 44.4%; Pred. No. 1.9e-19;
 Matches 91; Conservative 38; Mismatches 71; Indels 5; Gaps 1;
 QY 6 ILGYWKIKGIQVPTRLLEYLEEYEEHLYERDEG-----DKWRNKKFELGLEFPNLPY 60
 DB 3 ILGYWVRGLTIPRLLETTSSYEKKRYAMGDAPDYDRSOWLNKFKGLDFFNLPYL 62
 QY 61 IDGDVKLTOSMAITRYIADKHNNMLGCPKRAEISMLEGAVLDIRYGVSRVAYSKDFETL 120
 DB 63 IDGSRKITCSNAIMRYLARKHKLGCETEERIRADIVENQVMDNRMLIMLCYNPDFEQ 122
 QY 121 KVDFLSKPEMLKMFEDRLCHTKTYLNGDQIVTHPDMYLDALDVLVYMDPCLDAFPKLV 180
 DB 123 KPEFLKTIPEKNKLYSEFLGKRPWFAGDKVTYVDFLAYDILDQYHIFEPKCLDAFPN 182
 QY 181 EKKRIEAIPOIDKYLKSSKYIAMPL 205
 DB 183 FLARFELKKISAYMKSSRYLSTPI 207
 RESULT 10
 ID GTMUL_RABIT STANDARD; PRT; 217 AA.
 AC P46409;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE GLUTATHIONE S-TRANSFERASE MU 1 (EC 2.5.1.18) (GST MU 1) (GST CLASS-
 DE MU).
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=95251394; PubMed=7733673;
 RA Lee S.H., Lee S.H., Han J.S., Kim Y.S., Koh J.K.;
 RT "Cloning and expression of a cDNA for mu-class glutathione
 RT S-transferase from rabbit liver.";
 RL Arch. Biochem. Biophys. 318:424-429(1995).
 CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE = HX + R-S-GLUTATHIONE.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- TISSUE SPECIFICITY: WELL EXPRESSED IN RABBIT LIVER, BRAIN, AND
 CC KIDNEY.
 CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91239584; PubMed=2034681;
 RA Vorachek W.R., Pearson W.R., Rule G.S.;
 RT "Cloning, expression, and characterization of a class-mu glutathione
 transferase from human muscle, the product of the GST4 locus.";
 RT Proc. Natl. Acad. Sci. U.S.A. 88:4443-4447(1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS).
 RX MEDLINE=94238693; PubMed=8182750;
 RA Raghunathan S., Chandross R.J., Kretsinger R.H., Allison T.J.,
 RA Penington C.J., Rule G.S.;
 RT "Crystal structure of human class mu glutathione transferase GSTM2-2.
 Effects of lattice packing on conformational heterogeneity.";
 RT J. Mol. Biol. 238:815-832(1994).
 RL J. Mol. Biol. 238:815-832(1994).
 CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE = HX + R-S-GLUTATHIONE.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- TISSUE SPECIFICITY: THIS IS A MUSCLE ISOZYME.
 CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M63509; AAA60963.1; -;
 DR PIR: A39375; A39375.
 DR PDB: 1HNA; 31-JAN-94.
 DR PDB: 1HNB; 31-JAN-94.
 DR PDB: 1HNC; 31-JAN-94.
 DR PDB: 2GTU; 02-MAR-99.
 DR MTM: 138380; -;
 DR InterPro: IPR000521; GST.
 DR InterPro: IPR003081; GST_mu.
 DR Pfam: PF00043; GST; 1.
 DR PRINTS: PR01267; GSTRNSFRASEM.
 DR Transferase; Multigene family; 3D-structure.
 FT INIT_MET 0 0
 FT STRAND 2 5
 FT STRAND 7 7
 FT TURN 11 12
 FT HELIX 13 22
 FT TURN 23 24
 FT STRAND 27 29
 FT STRAND 32 32
 FT STRAND 35 35
 FT TURN 38 39
 FT STRAND 41 41
 FT HELIX 43 49
 FT TURN 50 52
 FT STRAND 61 64
 FT TURN 65 66
 FT STRAND 67 70
 FT HELIX 72 81
 FT TURN 82 84
 FT HELIX 90 113
 FT TURN 114 115
 FT TURN 117 118
 FT TURN 119 127
 FT HELIX 128 129
 FT TURN 130 141
 FT TURN 142 143
 FT TURN 146 146
 FT STRAND

FT TURN 147 147
 FT STRAND 148 149
 FT TURN 150 150
 FT STRAND 151 151
 FT HELIX 154 169
 FT HELIX 171 176
 FT HELIX 178 189
 FT TURN 191 197
 FT TURN 198 198
 FT TURN 214 215
 SQ SEQUENCE 217 AA; 25613 MW; 25603A909482CA39 CRC64;
 Query Match 10.9%; Score 473.5; DB 1; Length 217;
 Best Local Similarity 44.3%; Pred. No. 5e-19;
 Matches 94; Conservative 35; Mismatches 78; Indels 5; Gaps 1;
 QY 7 LGYWKIKGLVQPTRLLEYLEEYEEHLYERDEG-----DKWRNKKFELGLEFPNLPYYI 61
 DB 4 LGYWNIRGLAHSIRLLEYSYEEKKYTMGDAPDYDRSQWLNEKPKGLGLDFNLPYLI 63
 QY 62 DGDVKLTOSMAIIRYIADKHNMLGCGCPKERAETISMLECAVLDIRYGSRIAYSKDFTLK 121
 DB 64 DQTHKITQSNAILRYIARKHNLGCESEKEQIREIDILENQFMDSRMQLAKYCYDDPFELK 123
 QY 122 VDFLSKLPMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVLVMDPMCLDAFPLKVCVF 181
 DB 124 PEYLALEPEMLKLSQFCKQKOPWFLGDKLITFVDFLAYDVLERNQVFEPSCLDAFPLKDF 183
 QY 182 KKRIEAIPOIDKYIKSSKYIAMIPOGWOATFG 213
 DB 184 ISRPEGLEKISAYMKSSRFLPRPVTKMAVWG 215
 RESULT 13
 GTMU_CAVPO
 ID GTMU_CAVPO STANDARD; PRT; 217 AA.
 AC P16413;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE GLUTATHIONE S-TRANSFERASE B (EC 2.5.1.18) (GST B) (GST CLASS-MU).
 GN GSTM1.
 OS Cavia porcellus (Guinea pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Hystriognathi; Caviidae; Cavia.
 OX NCBI_TaxID=10141;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Liver;
 RX MEDLINE=90236961; PubMed=2332413;
 RA Kamei K., Oshino R., Hara S.;
 RT "Amino acid sequence of glutathione S-transferase b from guinea pig
 liver.";
 RL J. Biochem. 107:111-117(1990).
 CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE = HX + R-S-GLUTATHIONE.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY.
 DR PIR: JX0095; JX0095.
 DR HSSP: P04905; GGSV.
 DR InterPro: IPR000521; GST.
 DR InterPro: IPR003081; GST_mu.
 DR Pfam: PF00043; GST; 1.
 DR PRINTS: PR01267; GSTRNSFRASEM.
 DR Transferase; Multigene family.
 SQ SEQUENCE 217 AA; 25719 MW; D29F7951D4E9365E CRC64;

Query Match 10.9%; Score 473.5; DB 1; Length 217;
 Best Local Similarity 45.1%; Pred. No. 5e-19;

DR
DR PIR; S21908; S21908.
DR HSP; P04905; 5GST.
DR InterPro; IPR000521; GST.
DR InterPro; IPR003081; GST_mu.
DR Pfam; PF00045; GST; 1.
DR PRINIS; PR01267; GSTNRSFRASEM.
KW Transferase; Multigene family.

Tue Jan 15 14:12:07 2002

FT INIT MET 0 0 BY SIMILARITY.
SQ SEQUENCE 217 AA: 25558 MW: C59812A4A30B6375 CRC64;

Query Match 10.8%; Score 468.5; DB 1; Length 217;
Best Local Similarity 44.7%; Pred. No. 9.2e-19;
Matches 89; Conservative 37; Mismatches 68; Indels 5; Gaps 1;

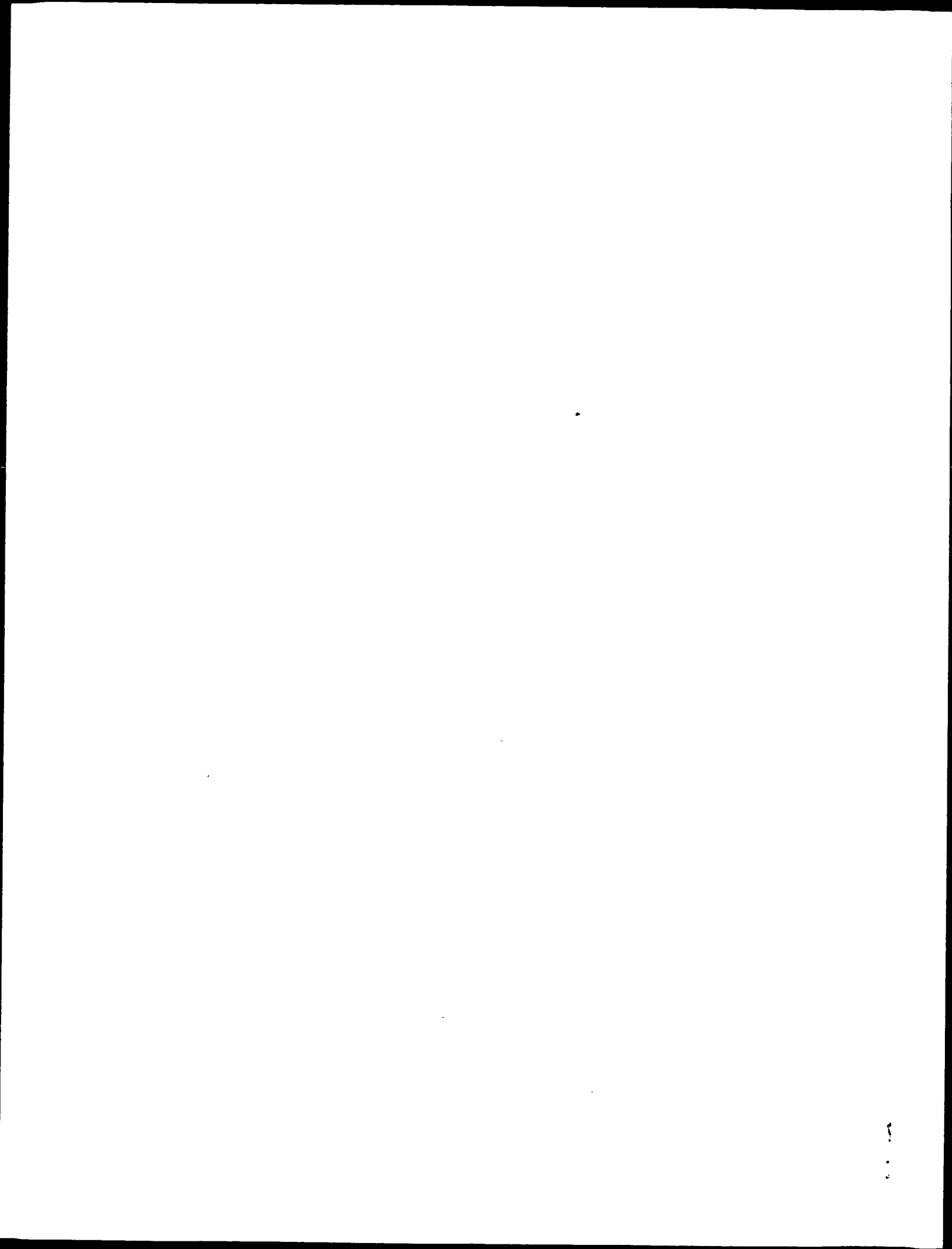
QY 7 LGYWKIKGLVOPTRLLLEYLEEKYEELHYERDEG-----DKWRNKKFELGLEFPNLPYI 61
Db 4 LGYWDIRGLAHAIIRLLLEVTDTSYEEKYTMGDAPNFRSQWLNKFKLGLDLPNLPYLI 63

QY 62 DGDVKLTQSMATIRYIADKHNMGLGCPKERAETISMLEGAVLDIRYGVSRVAYSKDFETLK 121
Db 64 DGSCHKITQSNAILRYIARKHDLGCGETEERIQDIIENQAMDTRMOLAMVCYSPDFEKKR 123

QY 122 VDFLSKLPEMKMFEDRLCHKTYLNGDHTVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCF 181
Db 124 PEYLEGLPEKMKLYSEFLGKRKSWFAGDKITYVDFLIYDVLQDHRIFAPKCLDAFPNPKDF 183

QY 182 KKRTEALPQIDKYLKSSKY 200
Db 184 LARFEGCLKISDYMKSSRF 202

Search completed: January 15, 2002, 14:07:00
Job time: 929 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:49:59 ; Search time 120.01 Seconds
(without alignments)
228.991 Million cell updates/sec

Title: US-09-724-296-4
Perfect score: 1979
Sequence: 1 DDHAPREFDCLDRPIWRG.....KRLTARKRRSRKEVEDEK 371

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: .522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	1979	100.0	371	21	AA144499 S. pombe delta228-
2	1979	100.0	600	21	AA144500 GST signal peptide
3	1979	100.0	828	21	AA144498 GST signal peptide
4	960	48.5	656	17	AAW01618 Neurospora crassa
5	910	46.0	626	21	AA144502 N. crassa delta228
6	910	46.0	626	21	AA144503 B. subtilis delta2
7	230.5	11.6	294	21	AA144505 D. radiodurans del
8	108.5	5.5	388	21	AA144507 Arabidopsis thalia
9	108.5	5.5	470	21	AA144509 Arabidopsis thalia
10	108.5	5.5	486	21	AA144510 Arabidopsis thalia
11	107.5	5.4	615	21	AA18697 Arabidopsis thalia

12	107.5	5.4	615	21	AA18697 Arabidopsis thalia
13	107.5	5.4	633	21	AA18696 Arabidopsis thalia
14	107.5	5.4	633	21	AA18696 Arabidopsis thalia
15	107.5	5.4	662	21	AA18695 Arabidopsis thalia
16	107.5	5.4	662	21	AA18695 Arabidopsis thalia
17	104	5.3	463	18	AAW34262 A beta subunit of
18	101	5.1	800	11	AA18695 Arabidopsis thalia
19	99	5.0	463	21	AA18695 Arabidopsis thalia
20	99	5.0	471	21	AA18695 Arabidopsis thalia
21	99	5.0	480	21	AA18695 Arabidopsis thalia
22	98.5	5.0	321	21	AA18695 Arabidopsis thalia
23	98.5	5.0	337	21	AA18695 Arabidopsis thalia
24	98.5	5.0	338	21	AA18695 Arabidopsis thalia
25	97.5	4.9	470	21	AA18695 Arabidopsis thalia
26	97.5	4.9	478	21	AA18695 Arabidopsis thalia
27	97.5	4.9	487	21	AA18695 Arabidopsis thalia
28	97	4.9	533	21	AA18695 Arabidopsis thalia
29	94	4.7	2110	21	AA18695 Arabidopsis thalia
30	93.5	4.7	2048	22	AA18695 Arabidopsis thalia
31	93	4.7	329	18	AA18695 Arabidopsis thalia
32	93	4.7	605	22	AA18695 Arabidopsis thalia
33	93	4.7	747	21	AA18695 Arabidopsis thalia
34	93	4.7	747	21	AA18695 Arabidopsis thalia
35	92.5	4.7	1059	21	AA18695 Arabidopsis thalia
36	91.5	4.6	653	19	AA18695 Arabidopsis thalia
37	91.5	4.6	1128	20	AA18695 Arabidopsis thalia
38	90.5	4.6	413	22	AA18695 Arabidopsis thalia
39	90	4.5	677	20	AA18695 Arabidopsis thalia
40	90	4.5	677	22	AA18695 Arabidopsis thalia
41	90	4.5	793	18	AA18695 Arabidopsis thalia
42	89.5	4.5	349	21	AA18695 Arabidopsis thalia
43	89.5	4.5	1780	22	AA18695 Arabidopsis thalia
44	89.5	4.5	3066	18	AA18695 Arabidopsis thalia
45	89	4.5	215	21	AA18695 Arabidopsis thalia

ALIGNMENTS

RESULT 1
AA144499
ID AA144499 standard; Protein; 371 AA.
XX
XX AA144499;
XX
XX 27-MAR-2000 (first entry)
DT
XX
XX S. pombe delta228-UV damage endonuclease.
DE
XX Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;
KW glutathione-S-transferase signal peptide; uvel+ gene product;
KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
KW abasic site; platinum diadduct; mismatched nucleotide pairing;
KW nucleotide alkylation; skin cancer.
XX
XX Schizosaccharomyces pombe.
XX
XX WO963828-A1.
XX
XX 16-DEC-1999.
PD
XX
XX 08-JUN-1999; 99WO-US12910.
PF
XX
XX 08-JUN-1998; 98US-0088521.
PR
XX
XX 18-MAY-1999; 99US-0134752.
PA
XX
XX (UYEM-) UNIV EMORY.
XX
XX Doetsch PW, Kaur B, Avery AM;
XX
XX WPI: 2000-116417/10.
DR
XX
XX N-PSDB: AA229859.

PT A new truncated ultraviolet damage endonuclease for treatment of skin
 XX cancers -
 XX
 PS Claim 13; Page 53; 133pp; English.
 XX
 CC The present sequence is delta228-UV damage endonuclease. Delta228-UVDE is
 CC a 288 amino acid deletion of the N-terminal of the S. pombe uvel+ gene
 CC product. This is expressed in frame with a GST leader sequence to
 CC generate a fusion protein. This provides stable endonuclease fragments
 CC for cleaving a double-stranded DNA molecule that has a distorted
 CC structure resulting from UV radiation damage, a photoproduct, an abasic
 CC site, mismatched nucleotide pairing, platinum diadduct, an intercalated
 CC molecule or alkylation of a nucleotide. Uvelp can be used in compositions
 CC for internal or topical application and as a therapeutic agent for skin
 CC cancers.
 XX
 SQ Sequence 371 AA;

Query Match 100.0%; Score 1979; DB 21; Length 371;
 Best Local Similarity 100.0%; Pred. No. 2.3e-187;
 Matches 371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DDHAPREMFDCLDKPIPWGRGLGYACLNILSRMKERVFCSTCRITTIQDGLSVKQL 60
 Db 1 ddhapremfdcldkpiwgrglgyacnlilrsmkervfcsrtcrittiqrdglesvkql 60
 QY 61 GTONVLDLILKLVENHNFGIHFMVSSDLFPFASHAKYGYTLEFAQSHLEEVGKLANKYN 120
 Db 61 gtqnvlldliklvewnhnfhgmrvssdlfpfashakgytlfefaqshleevgklankyn 120
 QY 121 HRLTMHPGGYTOIASPREVVVDASIRDLAYHDEILSRMKLNQKDAVLIHILGGTFEG 180
 Db 121 hrltmhpggytgiaprevvvdasirdlayhdeilsrmlneqinkdavlilhlggtfeg 180
 QY 181 KKETLDRKFNKYORLSDSVKARLVLENDVSVSDLLPLCOELNIPVLVDWHHNIVPG 240
 Db 181 kketldrfknqyrlsdsvkarlvlendvsvsqdllplcqelnipvlvdwhhnivpg 240
 QY 241 TLREGSLDMLPIPTIRETWTRKGITQKHYESADPTAISGMKRRASDRVDFPFCDDP 300
 Db 241 tlregslldmpliptiretwtrkgitqkhyesadptaisgmkrasdrvdfpfcddp 300
 QY 301 TMDLMEAKEKEQAVFELCRRYELQNPCCPLEIMGPEYDQTRDGYYPGAEKRLTARKR 360
 Db 301 tmdlmeakekeqavfelcrryelnppccpleimgpeydtqrdgyypgackrltarkr 360
 QY 361 SRKEEVEDEK 371
 Db 361 srkeeveedek 371

RESULT 2
 AA144500
 ID AA144500 standard; Protein; 600 AA.
 XX
 AC AA144500;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE GST signal peptide and delta228 S. pombe UVDE fusion protein.
 XX
 KW GST signal peptide; glutathione-S-transferase signal peptide;
 KW ultraviolet damage endonuclease; UVDE; delta228-UVDE;
 KW repair-deficient E. coli strain; UV irradiation; DNA damage;
 KW UV radiation damage; photoproduct; abasic site; platinum diadduct;
 KW mismatched nucleotide pairing; nucleotide alkylation;
 KW fusion protein; skin cancer.
 XX
 OS Schizosaccharomyces pombe.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers

FT Misc-difference 11 /note= "Encoded by aaaa"
 FT Misc-difference 85 /note= "Encoded by gtt"
 FT Misc-difference 147 /note= "Encoded by gtl"
 XX W09963828-A1.
 PN
 PD 16-DEC-1999.
 XX
 XX 08-JUN-1999; 99WO-US12910.
 XX
 XX 08-JUN-1998; 98US-0088521.
 PR
 PR 18-MAY-1999; 99US-0134752.
 XX
 XX (UYEM-) UNIV EMORY.
 PA
 XX Doetsch PW, Kaur B, Avery AM;
 PI
 XX WPI; 2000-1116417/10.
 DR
 DR N-PSDB; AA229860;
 XX
 XX A new truncated ultraviolet damage endonuclease for treatment of skin
 PT cancers -
 PT
 XX Disclosure; Page 56; 133pp; English.
 PS
 XX The present sequence is a fusion protein comprising the GST signal
 CC peptide and delta228 truncated ultraviolet damage endonuclease (UVDE)
 CC from S. pombe. Delta228-UVDE is a 288 amino acid deletion of the
 CC N-terminal of the S. pombe uvel+ gene product. This is expressed in frame
 CC with a GST leader sequence. Stable endonuclease fragments can be produced
 CC for cleaving a double-stranded DNA molecule that has a distorted
 CC structure resulting from UV radiation damage, a photoproduct, an abasic
 CC site, mismatched nucleotide pairing, platinum diadduct, an intercalated
 CC molecule or alkylation of a nucleotide. Uvelp can be used in compositions
 CC for internal or topical application and as a therapeutic agent for skin
 CC cancers.
 XX
 SQ Sequence 600 AA;
 Query Match 100.0%; Score 1979; DB 21; Length 600;
 Best Local Similarity 100.0%; Pred. No. 4.8e-187;
 Matches 371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DDHAPREMFDCLDKPIPWGRGLGYACLNILSRMKERVFCSTCRITTIQDGLSVKQL 60
 Db 230 ddhapremfdcldkpiwgrglgyacnlilrsmkervfcsrtcrittiqrdglesvkql 289
 QY 61 GTONVLDLILKLVENHNFGIHFMVSSDLFPFASHAKYGYTLEFAQSHLEEVGKLANKYN 120
 Db 290 gtqnvlldliklvewnhnfhgmrvssdlfpfashakgytlfefaqshleevgklankyn 349
 QY 121 HRLTMHPGGYTOIASPREVVVDASIRDLAYHDEILSRMKLNQKDAVLIHILGGTFEG 180
 Db 350 hrltmhpggytgiaprevvvdasirdlayhdeilsrmlneqinkdavlilhlggtfeg 409
 QY 181 KKETLDRKFNKYORLSDSVKARLVLENDVSVSDLLPLCOELNIPVLVDWHHNIVPG 240
 Db 410 kketldrfknqyrlsdsvkarlvlendvsvsqdllplcqelnipvlvdwhhnivpg 469
 QY 241 TLREGSLDMLPIPTIRETWTRKGITQKHYESADPTAISGMKRRASDRVDFPFCDDP 300
 Db 470 tlregslldmpliptiretwtrkgitqkhyesadptaisgmkrasdrvdfpfcddp 529
 QY 301 TMDLMEAKEKEQAVFELCRRYELQNPCCPLEIMGPEYDQTRDGYYPGAEKRLTARKR 360
 Db 530 tmdlmeakekeqavfelcrryelnppccpleimgpeydtqrdgyypgackrltarkr 589
 QY 361 SRKEEVEDEK 371
 XX

```

Query Match      100.0%; Score 1979; DB 21; Length 828;
Best Local Similarity 100.0%; Pred. No. 7.9e-187;
Matches 371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qv      1 DDAHPREMFDCDKPIPMWRGLGVACLNTILRSMKRVFCSTCRITTTIQRDGLSEVQQL 60

```


PN W09963828-A1.
 XX
 PD 16-DEC-1999.
 XX
 PF 08-JUN-1999; 99WO-US12910.
 XX
 PR 08-JUN-1998; 98US-0088521.
 PR 18-MAY-1999; 99US-0134752.
 XX
 PA (UYEM-) UNIV EMORY.
 XX
 PI Doetsch PW, Kaur B, Avery AM;
 XX WPI; 2000-116417/10.
 XX
 DR A new truncated ultraviolet damage endonuclease for treatment of skin
 PT cancers -
 PT
 PS Claim 16; Page 59; 133pp; English.
 XX
 CC The present sequence is delta228-UV damage endonuclease from B. subtilis.
 CC Delta228-UVDE is a 288 amino acid deletion of the N-terminal of the
 CC uvel+ gene product. This is expressed in frame with a GST leader sequence
 CC to generate a fusion protein. This provides stable endonuclease fragments
 CC for cleaving a double-stranded DNA molecule that has a distorted
 CC structure resulting from UV radiation damage, a photoproduct, an abasic
 CC site, mismatched nucleotide pairing, apurinium diadduct, an intercalated
 CC molecule or alkylation of a nucleotide. Uvelp can be used in compositions
 CC for internal or topical application and as a therapeutic agent for skin
 CC cancers.
 XX
 SQ Sequence 626 AA;
 Query Match 46.0%; Score 910; DB 21; Length 626;
 Best Local Similarity 45.1%; Pred. No. 3.7e-81;
 Matches 197; Conservative 51; Mismatches 91; Indels 98; Gaps 10;
 QY 15 PIPWRGLGVACLTILSRMKERVFCSTCRITTI----- 49
 Db 201 plpwkgrlgvacintyrnakppifssrtcrmasivdhrhplqfedepehlnknpkpsk 260
 QY 50 ---QRDGLSVKQLGTQNVLDLKLVEVNHNFHGFHFRVSSDLFPFASHAKYCYTL-EFA 105
 Db 261 epqdelghkfvqelglanardivkmclw-----fpfashvhygkiapfa 305
 QY 106 QSHLEEVGKLANKYNHRLTMHPGQYTOIASPREVVDSALRLAYHDEILSRMKLNEQLN 165
 Db 306 seviaeagrvaaelghrltthpgqftqgsprkevvesairdleyhdelisliklpeqqn 365
 QY 166 KDAVLIILHGGTFEGKXETLDRFRKNYORLSDSVKARLVLENDDVSVSVDLLPLCQELN 225
 Db 366 rdavmihmgggqfdgkaatlerfkrnyarlsqsknrlvlenddvgvtnhllpvceeln 425
 QY 226 IPLVDWHHINIV--PGLTREGSLDM--PLIPTIRETWKRGITOKOHYESADPTAIS 281
 Db 426 ipmvlghhnicfdpahlregtdidsdpklqeriantwkrkglkkmhysepcd-gavt 484
 QY 282 GMRRAHSDRVDFPPCDPTMDLMIEAKEQAVFELCRRYEL-----Q 325
 Db 485 prdrkrhrvmtlppcpdpmdlmieakkeqavfelmtfklpgfekindmvpvpyardde 544
 QY 326 NPCC-----PLEINGPEVDQTRDG-----YPPGAERLTA 356
 Db 545 nrpappvkapkkkgkrkrttdaeaaeeveepeeramgpnrvywpplgceewlkp 604
 QY 357 RKRSRK-----EEVEED 369
 Db 605 kkrvxxgkvpveede 621
 RESULT 7
 AAY44505

ID AAY44505 standard; Protein; 294 AA.
 XX
 AC AAY44505;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE D. radiodurans delta228-UV damage endonuclease.
 XX
 KW Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;
 KW glutathione-S-transferase signal peptide; uvel+ gene product;
 KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
 KW abasic site; apurinium diadduct; mismatched nucleotide pairing;
 KW nucleotide alkylation; skin cancer.
 XX
 OS Dienococcus radiodurans.
 XX
 PN W09963828-A1.
 XX
 PD 16-DEC-1999.
 XX
 PF 08-JUN-1999; 99WO-US12910.
 XX
 PR 08-JUN-1998; 98US-0088521.
 PR 18-MAY-1999; 99US-0134752.
 XX
 PA (UYEM-) UNIV EMORY.
 XX
 PI Doetsch PW, Kaur B, Avery AM;
 XX WPI; 2000-116417/10.
 XX
 DR A new truncated ultraviolet damage endonuclease for treatment of skin
 PT cancers -
 PT
 PS Claim 16; Page 60; 133pp; English.
 XX
 CC The present sequence is delta228-UV damage endonuclease from the
 CC D. radiodurans. Delta228-UVDE is a 288 amino acid deletion of the
 CC N-terminal of the uvel+ gene product. This is expressed in frame with a
 CC GST leader sequence to generate a fusion protein. This provides stable
 CC endonuclease fragments for cleaving a double-stranded DNA molecule that
 CC has a distorted structure resulting from UV radiation damage, a
 CC photoproduct, an abasic site, mismatched nucleotide pairing, apurinium
 CC diadduct, an intercalated molecule or alkylation of a nucleotide. Uvelp
 CC can be used in compositions for internal or topical application and as a
 CC therapeutic agent for skin cancers.
 XX
 SQ Sequence 294 AA;
 Query Match 11.6%; Score 230.5; DB 21; Length 294;
 Best Local Similarity 27.9%; Pred. No. 2.3e-14;
 Matches 88; Conservative 42; Mismatches 140; Indels 45; Gaps 10;
 QY 21 RLGYACILNLTLSMKERVFCSTCRITTIQRDGLSVKQLGTQNVLDLKLVEVNHNFHFI 80
 Db 1 qglvclvtgvpvfrftvltlsryalspaerea--kildlyssnktlrgaadycaahti 58
 QY 81 HFMVRSSDLFP---FASHAKYGYTLEFAQSHLEEVGKLANKYNHRLTMHPGQYTOIASPR 137
 Db 59 rlyrisslfpmlldlagddtgavithlapqlleaghaftdagvrlmhpeqfivlnsr 118
 QY 138 EVVVSQAIKRLAYHDEI-----LSRMKLNELQNLDAVLIIHLGGTFEGKXETLDRFRKNY 192
 Db 119 pevressvramsaaharvmadglartpwn-----lillhggkggrgae-----laali 166
 QY 193 QRLSDSVKARLVLENDDVSVSVDLLPLCQELNIPVLVDWHHINIVPGTLREGSLDMLP- 251
 Db 167 pdlpvrlrlglenderayspaellpiceatgtplvfadhhhv-----hdklpd 217
 QY 252 -LIPTIRETWKRGIT-----QKOHYESADPTAISGMRRAHSDRVDFPPCDPTM-D 303
 Db 218 qedpsvrewlratwqppewgvhls-----nglegpqdrhrshlladfpayadvpq 272

QY 304 LMIEAKEKEQAVFEL 318
: :||| ||:| :
Db 273 leevakgkeaaal 287

RESULT 8
AAG41350
ID AAG41350 standard; Protein: 388 AA.
XX
AC AAG41350;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 51435.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136031.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.


```
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150366.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      5.5%; Score 108.5; DB 21; Length 388;
Best Local Similarity 20.0%; Pred. No. 0.042;
Matches 69; Conservative 60; Mismatches 123; Indels 93; Gaps 16;

QY 27 LNTILSMKKEVFE-----CSRTRCITTTIQDGLSVKQIGTQNVLDLIKLVENH 76
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 54 mnsIargqkIpIIsaaqlphneIaaqlcr-----qagI---vkrlI---ektvdIle---dh 100
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 77 ---NFGIHFMRVSSDLFPFASHAKYGYTLEFAQSHLEEVG-----KLANK----- 118
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
```

```
Db 101 gednfaivfaamgvnm-----etaqfkfrdfceangsmervtllflnlandptieri 150
QY 119 YNHRLTMHPCQYQTASPREVV-----DSAIRDL-AYHDEILSRMKLNEQLNKDAV 169
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 151 itprialttaeylayecgkhviltmdssyadaliressaareevpgrgypgymytdlia 210
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 170 LIIHLGGTFEaKFTLDRFRKNYQRLSDSVKARLVLENDVSVQDILLPLCOELNIPLV 229
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 211 tiyeragriegrksitql-----pltmpnddithptpditgyitegqiylid 258
QY 230 LDWHHHNIVPGTLREGSLDMLPIPTIRETWTRKGTOKQH-----YSESADPTAISG 282
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 259 rqlhnrqiyp-----pinvlpslrmlksaigemtrkdhdsnqnyanyaigkdvqa 312
QY 283 MKRRASHDRVDFPPCDPTMDLM-IEAKEKEQAVFPLCRRYELQN 326
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 313 mkavvgeaal-----ssedllylefldkferkfvmqgaydtrn 350

RESULT 9
AAG41349
ID AAG41349 standard; Protein; 470 AA.
XX
AC AAG41349;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 51434.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
```

PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139432.
PR 18-JUN-1999; 99US-0139434.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142330.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 29-SEP-1999; 99US-0157117.
PR 04-OCT-1999; 99US-0157753.
PR 05-OCT-1999; 99US-0157865.
PR 06-OCT-1999; 99US-0158029.
PR 07-OCT-1999; 99US-0158232.
PR 08-OCT-1999; 99US-0158369.
PR 12-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159299.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.

```
PR 28-OCT-1999; 99US-01611993.
PR 29-OCT-1999; 99US-0162142.

Query Match          5.5%; Score 108.5; DB 21; Length 470;
Best Local Similarity 20.0%; Pred. No. 0.056;
Matches 69; Conservative 60; Mismatches 123; Indels 93; Gaps 16;

QY 27 LNTILSRKERVF-----CSRTCRITTIQRDGLSVKQLGTQNVLDLIKLVENNH 76
Db 136 mnsiargqkplfsaaglpheiaagqcr-----qagl--vkrl--ektdvll-----dh 182
QY 77 ---NFGIHFRVSDLPFPFASHAKYGYTLEFAQSHLEEVG-----KLANK----- 118
Db 183 gednfaivfaamgvnm-----etaqfkrdfceengsmervtlfifinlandptieri 232
QY 119 YNURLTMHPQOYTOIASPREVV-----DSAIRDL-AYHDEILSRMKLNQOLMKDAV 169
Db 233 itprialttaeylayecgkhviltmdssyadalrevaareevpgrrgpygmytdia 292
QY 170 LIHLGTFEGKKTLDLFRKNYQRLSDSVKARLVLENDVSWQDILLPLCOELNIPLV 229
Db 293 tiyeragregkrqstqi-----piltmpnddithptdltgyitedgqiylid 340
QY 230 LDWHHNIIVPGTLREGSLDLMLPIPTIRETWKRGITQKH-----YESADPTAISG 282
Db 341 rqlhnrqiyp-----pinvlpsalsrlmksaigegmtrkdhdsvsnqlyanyaigkdvqa 394
QY 283 MKRRASDRVDFPDCPTMDLM-IEAKEQAVFELCRRYELON 326
Db 395 mkavvgeaal-----ssedillylefldkferkfvmqgaydtrn 432

RESULT 10
AAG41348
ID AAG41348 standard; Protein; 486 AA.
XX
AC AAG41348;
DT
DE 18-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 51433.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
```


PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138340.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.


```
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145318.
PR 27-JUL-1999; 99US-0145319.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
```

```
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.4%; Score 107.5; DB 21; Length 615;
Best Local Similarity 22.6%; Pred. No. 0.1;
Matches 61; Conservative 34; Mismatches 98; Indels 77; Gaps 11;

QY 48 TIORDGLESVKOL-----GTQNVLDLIKLVENNNHFGIHFMRVSSDLFPFASHAKYGYT 101
Db 344 timeqglrdlqleqdlvfgdagrkadvkf1stnh-----iish----- 382

QY 102 LEFAOSHLEEVGKLANKYNHRLTMHPGQYQIASPREVVVDSAIRDLAYHDEILSRMK1N 161
Db 383 ----eskl-----rllimivaaipkpkfegekgrkmmel 411

QY 162 EQLNKDAVLIIH----LGGTF-EGKKETLDRFRKNYQRLSDSVKARLVLENDVSNVOD 216
Db 412 akslgddvvavnmrllgpvhtecksttgsfp1kfdv1kktkraardrvgetqtwqlsr 471

QY 217 LLPLQCEL-----NPLVLDMHHHNIIVPGTLREGSL--DLMPILPTIRE--TWTRKGI 265
Db 472 fypiveelvek1sknhlp-kqdybcmekpctfysgs1spsasvplphsrrtptwarh1 530

QY 266 TOKQHYSESADPTA1SGMKRRRAHSRVDVF 295
Db 531 sddgyfsvdsvlgrassgfrk--gqri1vf 558
```

RESULT 13

AAG18696

ID AAG18696 standard; Protein; 633 AA.

XX AAG18696;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 20208.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.

OS Arabidopsis thaliana.

XX EPI033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 18-MAY-1999; 99US-0134370.

PR 19-MAY-1999; 99US-0134768.

PR 20-MAY-1999; 99US-0134941.

PR 21-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139452.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139859.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.

PR 05-AUG-1999; 99US-0147192.

PR 06-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.

PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.

PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.

PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.

PR 17-AUG-1999; 99US-0149175.

PR 18-AUG-1999; 99US-0149426.

PR 20-AUG-1999; 99US-0149722.

PR 20-AUG-1999; 99US-0149723.

PR 20-AUG-1999; 99US-0149929.

PR 23-AUG-1999; 99US-0149902.

PR 23-AUG-1999; 99US-0149930.

PR 25-AUG-1999; 99US-0150566.

PR 26-AUG-1999; 99US-0150884.

PR 27-AUG-1999; 99US-0151065.

PR 27-AUG-1999; 99US-0151066.

PR 27-AUG-1999; 99US-0151080.

PR 30-AUG-1999; 99US-0151303.

PR 31-AUG-1999; 99US-0151438.


```
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 21-JUN-1999; 99US-0139899.
PR 21-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 23-JUN-1999; 99US-0140355.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142134.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147935.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.

PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.4%; Score 107.5; DB 21; Length 633;
Best Local Similarity 22.8%; Pred. No. 0.11; 98; Indels 77; Gaps 11;
Matches 61; Conservative 34; Mismatches 34;
QY 48 TQRDGLSVKQL-----GTQNVLDLKLVEWNHNFHGMVRVSSDLFPFASHAKYGYT 101
Db 362 timeggldlqleqdlvfgdagrkdvikflstnh-----ilsh----- 400
QY 102 LEFAQSHLEEVGKLANKYNHRLTMHPQYQIASPREVVVDSAIRDAYHDEILSRMKN 161
Db 401 -----eskl-----tlimivaalypkkfeyekgrkmmel 429
QY 162 EOLNKDAVLIH-----LGTF-EGKKTETDRFRKNYORLSDSVKARLVLENDVWSVQD 216
Db 430 akisgddvavnmrligpvhtecksttsfplkfdviktkaarrdrvgetqtqlsr 489
QY 217 LLLPLCQEL-----NIFLVLDWHHNNHNVPGTLEEGSL--DLMPLIPTIRE--TWTRKGI 265
```

Db	490	fypiveelvelksghlp-kqdybcmepkptfygslpsaspvlphsrrtptwarhl	548
QY	266	TQKHYSSEADPTAISGMKRAHSDRVDFD	295
Db	549	sddgyfdsdsvlgrassgfkrr-gqrifvf	576
RESULT 15			
AAG18695			
ID	AAG18695 standard; Protein; 662 AA.		
XX			
AC	AAG18695;		
XX			
DT	17-OCT-2000 (first entry)		
XX			
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 20207.		
XX			
KW	Protein identification; signal transduction pathway; metabolic pathway;		
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		
KW	termination sequence.		
XX			
OS	Arabidopsis thaliana.		
XX			
PN	EP1033405-A2.		
XX			
PD	06-SEP-2000.		
XX			
PF	25-FEB-2000; 2000EP-0301439.		
XX			
PR	25-FEB-1999;	99US-0121825.	16-JUN-1999;
PR	05-MAR-1999;	99US-0123180.	16-JUN-1999;
PR	09-MAR-1999;	99US-0123548.	17-JUN-1999;
PR	23-MAR-1999;	99US-0125788.	18-JUN-1999;
PR	25-MAR-1999;	99US-0126264.	18-JUN-1999;
PR	29-MAR-1999;	99US-0126785.	18-JUN-1999;
PR	29-MAR-1999;	99US-0127462.	18-JUN-1999;
PR	06-APR-1999;	99US-0128234.	18-JUN-1999;
PR	08-APR-1999;	99US-0128714.	18-JUN-1999;
PR	16-APR-1999;	99US-0129845.	18-JUN-1999;
PR	19-APR-1999;	99US-0130077.	18-JUN-1999;
PR	21-APR-1999;	99US-0130449.	18-JUN-1999;
PR	23-APR-1999;	99US-0130510.	18-JUN-1999;
PR	28-APR-1999;	99US-0130891.	18-JUN-1999;
PR	30-APR-1999;	99US-0131449.	18-JUN-1999;
PR	30-APR-1999;	99US-0132048.	18-JUN-1999;
PR	04-MAY-1999;	99US-0132407.	18-JUN-1999;
PR	03-MAY-1999;	99US-0132484.	18-JUN-1999;
PR	06-MAY-1999;	99US-0132485.	18-JUN-1999;
PR	06-MAY-1999;	99US-0132486.	18-JUN-1999;
PR	07-MAY-1999;	99US-0132863.	18-JUN-1999;
PR	11-MAY-1999;	99US-0134256.	18-JUN-1999;
PR	14-MAY-1999;	99US-0134218.	18-JUN-1999;
PR	14-MAY-1999;	99US-0134219.	18-JUN-1999;
PR	14-MAY-1999;	99US-0134221.	18-JUN-1999;
PR	18-MAY-1999;	99US-0134370.	18-JUN-1999;
PR	19-MAY-1999;	99US-0134768.	18-JUN-1999;
PR	20-MAY-1999;	99US-0134941.	18-JUN-1999;
PR	21-MAY-1999;	99US-0135124.	18-JUN-1999;
PR	24-MAY-1999;	99US-0135353.	18-JUN-1999;
PR	25-MAY-1999;	99US-0135629.	18-JUN-1999;
PR	27-MAY-1999;	99US-0136021.	18-JUN-1999;
PR	28-MAY-1999;	99US-0136392.	18-JUN-1999;
PR	01-JUN-1999;	99US-0136782.	18-JUN-1999;
PR	03-JUN-1999;	99US-0137222.	18-JUN-1999;
PR	04-JUN-1999;	99US-0137528.	18-JUN-1999;
PR	07-JUN-1999;	99US-0137502.	18-JUN-1999;
PR	08-JUN-1999;	99US-0137724.	18-JUN-1999;
PR	10-JUN-1999;	99US-0138094.	18-JUN-1999;
PR	10-JUN-1999;	99US-0138540.	18-JUN-1999;
PR	14-JUN-1999;	99US-0138847.	18-JUN-1999;
PR	14-JUN-1999;	99US-0139119.	18-JUN-1999;

PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.4%; Score 107.5; DB 21; Length 662;
Best Local Similarity 22.6%; Pred. No. 0.12;
Matches 61; Conservative 34; Mismatches 98; Indels 77; Gaps 11;

QY 48 TIRDCLESVKOL-----GTQNVLDLIKLVENHNFCIHFMRVSSDLFPFASHAKYGT 101
Db 391 timeqglrdlgqleqdvfgdagrvdkvfkisthh-----lsh----- 429

QY 102 LEFAQSHLEEVGKLANKYNHRLTMHFGQYQTIASPREVVVDSAIROLAYVHDEILSRMKLN 161
Db 430 -----eskl-----rlmivaaiypkkfegekgrkmmel 458
QY 162 EQLNKDAVLIH-----LGTF-EGKKEITLDRFRKNYQRLSDSVKARLVLENDVDSVSD 216
Db 459 aklsgddvavnnmrllgpyhteckkstgspfkldvltkraarrdrvgetqtwqjsr 518
QY 217 LLPLCOEL-----NIPLVLDWHHHNIVPGTLREGSL--DLMPLIPTIRE--TWTRKGI 265
Db 519 fypiveelveklskghlp-kqdyppcmnekpftfysgspspsasplphsrrtptwarthl 577
QY 266 TOKQHYSESADPTAISGMKRRRAHSRVDFF 295
Db 578 sddgyfsdsvlgrassgfkrrk--gqrifvf 605

Search completed: January 15, 2002, 13:50:00
Job time: 289 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:50:15 ; Search time 120.01 seconds
(without alignments)
181.464 Million cell updates/sec

Title: US-09-724-296-39

Perfect score: 1527

Sequence: 1 QLGVLCTGPEVRFRTVT.....VEAKGKEAIAALRLMAPFK 294

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_1101.*
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1527	100.0	294	21	AAV44505
2	255.5	16.7	656	17	AAW01618
3	239.5	15.7	626	21	AAV44502
4	239.5	15.7	626	21	AAV44503
5	230.5	15.1	371	21	AAV44499
6	230.5	15.1	600	21	AAV44500
7	230.5	15.1	828	21	AAV44498
8	94.5	6.2	270	14	AAW42817
9	93	6.1	199	19	AAW48651
10	92.5	6.1	915	22	AAE00207
11	90	5.9	497	22	AAG90354

12	90	5.9	497	22	AAV79113
13	90	5.9	1421	21	AAV58573
14	90	5.9	1891	19	AAW52844
15	90	5.9	3413	19	AAW52849
16	89	5.8	6797	22	AAW31558
17	88	5.8	537	18	AAW37045
18	86	5.6	1110	22	AAW81214
19	85.5	5.6	653	17	AAW98903
20	84.5	5.5	613	22	AAU04887
21	84.5	5.5	616	22	AAW39256
22	84.5	5.5	646	22	AAW41042
23	84	5.5	403	20	AAW82682
24	83	5.4	1291	20	AAW16101
25	83	5.4	1891	18	AAW23720
26	83	5.4	1391	18	AAW22610
27	82.5	5.4	1160	21	AAV52036
28	82.5	5.4	1160	21	AAV51665
29	82.5	5.4	1160	22	AAW98931
30	82.5	5.4	3567	14	AAW44431
31	82	5.4	404	10	AAW93216
32	82	5.4	1477	20	AAV19981
33	82	5.4	1494	20	AAV19980
34	81.5	5.3	198	22	AAW20162
35	81.5	5.3	465	22	AAW82580
36	81.5	5.3	5588	20	AAV39301
37	81.5	5.3	5588	22	AAW70969
38	81	5.3	506	9	AAW80802
39	81	5.3	521	9	AAW80807
40	81	5.3	607	18	AAW22357
41	80.5	5.3	186	21	AAW32418
42	80.5	5.3	187	21	AAW32417
43	80.5	5.3	221	21	AAW18250
44	80.5	5.3	221	21	AAW32416
45	80.5	5.3	249	21	AAW18249

ALIGNMENTS

RESULT 1

AAV44505

ID AAV44505 standard; Protein: 294 AA.

AC AAV44505;

XX 27-MAR-2000 (first entry)

DT D. radiodurans delta228-UV damage endonuclease.

DE Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;

KW glutathione-S-transferase signal peptide; uvex-gene product;

KW UV irradiation; DNA damage; UV radiation damage; photoproduct;

KW abasic site; apurinic diaduct; mismatched nucleotide pairing;

KW nucleotide alkylation; skin cancer.

XX D. radiodurans radiodurans.

OS D. radiodurans radiodurans.

XX WO9963828-A1.

PN 16-DEC-1999.

PD 08-JUN-1999; 99WO-US12910.

PF 08-JUN-1998; 98US-0088521.

PR 18-MAY-1999; 99US-0134752.

XX (UYEM-) UNIV EMORY.

XX Doetsch PW, Kaur B, Avery AM;

XX WPI; 2000-116417/10.

XX A new truncated ultraviolet damage endonuclease for treatment of skin

PT

PT cancers -
 XX Claim 16; Page 60; 133pp; English.
 XX
 CC The present sequence is delta228-UV damage endonuclease from
 CC D. radiodurans. Delta228-UVDE is a 288 amino acid deletion of the
 CC N-terminal of the uvel+ gene product. This is expressed in frame with a
 CC GST leader sequence to generate a fusion protein. This provides stable
 CC endonuclease fragments for cleaving a double-stranded DNA molecule that
 CC has a distorted structure resulting from UV radiation damage, a
 CC photoproduct, an abasic site, mismatched nucleotide pairing, a platinum
 CC diadduct, an intercalated molecule or alkylation of a nucleotide. Uvelp
 CC can be used in compositions for internal or topical application and as a
 CC therapeutic agent for skin cancers.
 XX
 XX Sequence 294 AA;

Query Match 100.0%; Score 1527; DB 21; Length 294;
 Best Local Similarity 100.0%; Pred. No. 3e-159;
 Matches 294; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLCGLVCTVGPEVRPTVTLRYRALSPAEREAKLLDLYSSNIKTLRGAADYCAAHDIRL 60
 DB 1 qglvctvgpevrftvltlsryalspaereaklldlyssniktlrqaadycaahdirl 60

QY 61 YRLSSSLFPMLDLAGDDTGA AVLTHLAPOLLEAGHAFTDAGVRLLMHPPEQFIVLNSDRPE 120
 DB 61 yrlssslfpmldlagdgtgaavtlhapqlleaghaftdagvrlmhppeqfivlnsdrepe 120

QY 121 VRESSVRAMSAHARVMDGLGLARTPNWLLLLHGGKGGRAELAAALIPDLPVRLRLGLE 180
 DB 121 vressvrmsaharvmdglglartpnwllllhggkgrgaelaalipdpvrrlrlgle 180

QY 181 NDERAYSPAELLIPCEATGTPLVFDAAHHVHDKLPDQEDPSVREWVLRARATWOPPEWQ 240
 DB 181 nderayspaellipceatgtplvfdahhhvhdklpdqedpsvrevvlaratwqppewq 240

QY 241 VVHLSNGIEGPDORHSHLIADFPAYADVPQIEVEAKGKEEAIAALRLMAPFK 294
 DB 241 vvhlslngiegpqdrhshliadfpayadvpqieveakgkeeaiaalrlmapfk 294

RESULT 2
 AAW01618
 ID AAW01618 standard; Protein; 656 AA.
 XX
 AC AAW01618;
 XX
 DT 20-AUG-1997 (first entry)
 XX
 DE Neurospora crassa DNA repair enzyme.
 XX
 KW DNA repair enzyme; cyclobutane-type dimer; (6-4)-bound product;
 KW exposure; UV light; ultraviolet; suntan cream; prevention; treatment;
 KW skin cancer.
 XX
 OS Neurospora crassa.
 XX
 PN JP08266275-A.
 XX
 PD 15-OCT-1996.
 XX
 PF 29-MAR-1995; 95JP-0094137.
 XX
 PR 29-MAR-1995; 95JP-0094137.
 XX
 PA (SHIS) SHISEIDO CO LTD.
 XX
 DR WPI; 1996-512663/51.
 DR N-PSDB; AAT58286.
 XX
 PT DNA repair enzyme and related DNA - specifically recognises

PT cyclobutane-type dimer and (6-4)-bound product induced by exposure
 PT to UV light
 XX
 PS Disclosure; Page 10-13; 18pp; Japanese.
 XX
 CC The present sequence shows a DNA repair enzyme which specifically
 CC recognises a cyclobutane-type dimer and (6-4)-bound product induced,
 CC respectively from TT and TC sequences in DNA, by exposure to UV light.
 CC The enzyme cleaves these sequences at the 5'-side. The DNA repair enzyme
 CC can be used in, e.g. suntan cream for prevention and treatment of skin
 CC cancer. Conventional endonuclease acts on either the cyclobutane-type
 CC dimer or the (6-4)-bound product in damaged DNA, while the enzyme
 CC of the invention can act on both of them.
 XX
 XX Sequence 656 AA;

Query Match 16.7%; Score 255.5; DB 17; Length 656;
 Best Local Similarity 31.8%; Pred. No. 5.3e-19;
 Matches 84; Conservative 38; Mismatches 103; Indels 39; Gaps 13;

QY 56 HDRLRYRLSSSLFPMLDLAGDDTCAAVLTHLAPQLL-BAGHAFTDAGVRLLMHPPEQFIVL 114
 DB 292 ygirflrlsemip---fashpvhyklapfasevtaeagrvaaelghrltthpgqftql 348

QY 115 NSDRPEVRESSVRAMSAHARVMDGLGLARTPNW---LLLHGGKGGRAELAAALIT--- 166
 DB 349 gspkrvevsaardleyhdei---lsllkpeqnrdaavmlih-mggqfgdkaatlerrf 403

QY 167 ---PDLDPVRLRLGLENDERAYSPAELLIPCEATGTPLVFDAAHHV---HDKL 215
 DB 404 knyarlsgckrlrvlndddvgwtvhdllpvcceelnipmvdvhhnfcfdpahiregt 463

QY 216 PDQEDPSVREWVLRARATWOPPE-WQVHLSNGIEG---PQDRK-HSHLIADFPAYADV 270
 DB 464 ldsdpklqe---riantwkrkqikkmhysepcdgavtprhrrkhrprvmtlppcpddm 520

QY 271 PQIEVEAKGKEEAIAALRLMAPFK 294
 DB 521 -dimieakdkeqav--felartfk 541

RESULT 3
 AAY44502
 ID AAY44502 standard; Protein; 626 AA.
 XX
 AC AAY44502;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE N. crassa delta228-UV damage endonuclease.
 XX
 KW delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;
 KW glutathione-S-transferase signal peptide; uvel+ gene product;
 KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
 KW abasic site; platinum diadduct; mismatched nucleotide pairing;
 KW nucleotide alkylation; skin cancer.
 XX
 OS Neurospora crassa.
 XX
 PN WO9963828-A1.
 XX
 PD 16-DEC-1999.
 XX
 PF 08-JUN-1999; 99WO-US12910.
 XX
 PR 08-JUN-1998; 98US-0088521.
 PR 18-MAY-1999; 99US-0134752.
 XX
 PA (UYEW-) UNIV EMORY.
 XX
 PI Doetsch PW, Kaur B, Avery AM;
 XX

DR WPI; 2000-116417/10.
XX A new truncated ultraviolet damage endonuclease for treatment of skin
PT cancers -
PT
XX
PS Claim 16; Page 59; 133pp; English.
XX
CC The present sequence is delta228-UV damage endonuclease from B. subtilis.
CC Delta228-UVDE is a 288 amino acid deletion of the N-terminal of the
CC uvel+ gene product. This is expressed in frame with a GST leader sequence
CC to generate a fusion protein. This provides stable endonuclease fragments
CC for cleaving a double-stranded DNA molecule that has a distorted
CC structure resulting from UV radiation damage, a photoproduct, an abasic
CC site, mismatched nucleotide pairing, platinum diadduct, an intercalated
CC molecule or alkylation of a nucleotide. Uvelp can be used in compositions
CC for internal or topical application and as a therapeutic agent for skin
CC cancers.
XX
SQ Sequence 626 AA;

Query Match 15.7%; Score 239.5; DB 21; Length 626;
Best Local Similarity 32.5%; Pred. No. 2.8e-17;
Matches 77; Conservative 34; Mismatches 87; Indels 39; Gaps 12;
QY 86 LAP---QLLEAGHAFTDAGVRLMHPEQFIVLNSDRPEVRESSVRAMSAHARYMDGGL 141
DB 301 lapfasevlaeagrvaaelghrlthpqqftqlgsprkeevvesairdleyhdel---lsl 357
QY 142 ARTPNW-----LLLLHGGKGGRAELAAII-----PDLPPVRLRLGLENDERAYSPA 189
DB 358 lkipeqgnrdavmliih--mgggfgdkaatlfrknyarlsgscknrivlendvgwtvh 415
QY 190 ELLPICEATGTPLVFDAAHHVV-----HDKLPQEDPSVREWVLRARATWQPPE-WQV 241
DB 416 dlipvecelnlpvldyhhnfcfpahrlregtldisdpkqe---riantwkrkgikqk 472
QY 242 VHLNSGTEG---PDQRR-HSHLIADPPSAYADVPQIEVEAKGKEBAIAALRLMAPFK 294
DB 473 mhysepcdgavtprdrkrhrprvmtlppcpddm-dlmieakdkeqav--feimrtfk 526

RESULT 5

AAAY44503
ID AAY44499 standard; Protein: 371 AA.
XX
AC AAY44499;
XX
DT 27-MAR-2000 (first entry)
XX
DE S. pombe delta228-UV damage endonuclease.
XX
KW Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;
KW glutathione-S-transferase signal peptide; uvel+ gene product;
KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
KW abasic site; platinum diadduct; mismatched nucleotide pairing;
KW nucleotide alkylation; skin cancer.
XX
OS Schizosaccharomyces pombe.
XX
PN WO9963828-A1.
XX
PD 16-DEC-1999.
XX
PF 08-JUN-1999; 99WO-US12910.
XX
PR 08-JUN-1998; 98US-0088521.
PR 18-MAY-1999; 99US-0134752.
XX
PA (UYEM-) UNIV EMORY.
XX
PI Doetsch PW, Kaur B, Avery AM;
XX

DR WPI; 2000-116417/10.
XX A new truncated ultraviolet damage endonuclease for treatment of skin
PT cancers -
PT
XX
PS Claim 16; Page 59; 133pp; English.
XX
CC The present sequence is delta228-UV damage endonuclease from N. crassa.
CC Delta228-UVDE is a 288 amino acid deletion of the N-terminal of the
CC uvel+ gene product. This is expressed in frame with a GST leader sequence
CC to generate a fusion protein. This provides stable endonuclease fragments
CC for cleaving a double-stranded DNA molecule that has a distorted
CC structure resulting from UV radiation damage, a photoproduct, an abasic
CC site, mismatched nucleotide pairing, platinum diadduct, an intercalated
CC molecule or alkylation of a nucleotide. Uvelp can be used in compositions
CC for internal or topical application and as a therapeutic agent for skin
CC cancers.
XX
SQ Sequence 626 AA;

Query Match 15.7%; Score 239.5; DB 21; Length 626;
Best Local Similarity 32.5%; Pred. No. 2.8e-17;
Matches 77; Conservative 34; Mismatches 87; Indels 39; Gaps 12;
QY 86 LAP---QLLEAGHAFTDAGVRLMHPEQFIVLNSDRPEVRESSVRAMSAHARYMDGGL 141
DB 301 lapfasevlaeagrvaaelghrlthpqqftqlgsprkeevvesairdleyhdel---lsl 357
QY 142 ARTPNW-----LLLLHGGKGGRAELAAII-----PDLPPVRLRLGLENDERAYSPA 189
DB 358 lkipeqgnrdavmliih--mgggfgdkaatlfrknyarlsgscknrivlendvgwtvh 415
QY 190 ELLPICEATGTPLVFDAAHHVV-----HDKLPQEDPSVREWVLRARATWQPPE-WQV 241
DB 416 dlipvecelnlpvldyhhnfcfpahrlregtldisdpkqe---riantwkrkgikqk 472
QY 242 VHLNSGTEG---PDQRR-HSHLIADPPSAYADVPQIEVEAKGKEBAIAALRLMAPFK 294
DB 473 mhysepcdgavtprdrkrhrprvmtlppcpddm-dlmieakdkeqav--feimrtfk 526

RESULT 4

AAAY44503
ID AAY44503 standard; Protein: 626 AA.
XX
AC AAY44503;
XX
DT 27-MAR-2000 (first entry)
XX
DE B. subtilis delta228-UV damage endonuclease.
XX
KW Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;
KW glutathione-S-transferase signal peptide; uvel+ gene product;
KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
KW abasic site; platinum diadduct; mismatched nucleotide pairing;
KW nucleotide alkylation; skin cancer.
XX
OS Bacillus subtilis.
XX
PN WO9963828-A1.
XX
PD 16-DEC-1999.
XX
PF 08-JUN-1999; 99WO-US12910.
XX
PR 08-JUN-1998; 98US-0088521.
PR 18-MAY-1999; 99US-0134752.
XX
PA (UYEM-) UNIV EMORY.
XX
PI Doetsch PW, Kaur B, Avery AM;
XX

QY	1	OLGLVCLTVGPEVFRFTVLRSRYRALSPAEREA--KLLDLYSSNIKTLRGAADYCAAHDI	58
	:	: :	:
	:	: :	:
Db	478	rlygacnlrlrmkervfcstcrttlqrdgliesvkigtqvldklilvevnhnfgl	537
	:	: :	:
QY	59	RLYRSSFLPMLDLAGDDTGAAVLTHLPQLLEAGHAFDTAGVRLLMHPQFIVLSNDR	118
	:	: :	:

Query Match 6.2%; Score 94.5; DB 14; Length 270;
Best Local Similarity 27.5%; Pred. No. 0.068;
Matches 70; Conservative 29; Mismatches 67; Indels 89; Gaps 17;
QY 22 RYRALSPATREAKLLDLYSSNKTIRGAADYCAA--HDIRLYRLSS--SLF--PMLIDLAG 75

from Coryneform bacterium, and identifying a homologue of a gene derived from coryneform bacterium. Coryneform bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a protein described in the exemplification of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office.

XX	Sequence	497 AA;
SO		

Query Match 5.9%; Score 90; DB 22; Length 497;
Best Local Similarity 23.5%; Pred. No. 0.54;
Matches 72; Conservative 31; Mismatches 90; Indels 114; Gaps 14;

20	LSYRYALSFAERAKLLDLYS	-----SNIKTLRGAAADYCAAGHLL	01
	: : : : : : : :	: : : : : : : :	
68	lkwsaltgaertgylikateiesrsealatlntrengspisetrgaasnaag	-----if	123
	: : : : : : : :	: : : : : : : :	
62	RLSSLPMLDLAGDB-----TGAA-----	-VILTHLAPOLLE	92
	: : : : : : : :	: : : : : : : :	
124	ryfatlapwid--gedirfpagsaesivdkdpgvcaliapwnfpinlvviklapall-	180	
	: : : : : : : :	: : : : : : : :	
93	AGHAFTDAGVRLIMHPE-----QFIVLNSDRPPEVRESSYRAMSAHARVMDGLGLARTP	145	
	: : : : : : : :	: : : : : : : :	
181	-----agctvliikpasptpIsirfiteaagvpagvvnltgsgrfgd--alvrhp	231	
	: : : : : : : :	: : : : : : : :	
146	WNLLLLHGKGGKGAELAAALIPDPVPLRLGLENDERAYSPAELIIPICETATPLVFD	205	
	: : : : : : : :	: : : : : : : :	
232	gydkvafgtspvdkkilaacgellrpvtlrlg-----gkssailp-----d	274	
	: : : : : : : :	: : : : : : : :	
206	AHHHVHDKLPDQEDSPVREWLRAKATQPPPEWQVVLHNSGIEGPDQRHSHLIADEPPS	265	
	: : : : : : : :	: : : : : : : :	
275	admsvlstrl-----irscm--rnt-----gqcyistrliaapssr-----	308	
	: : : : : : : :	: : : : : : : :	
266	AYADVPQ	272	
	: : : :	: : : :	
309	-vaevvq	314	
	: : : :	: : : :	

RESULT 12

AAB79113
AAB79113 standard. Protein: 497 AA.

XX AAB79113;
AC

XX
30-APR-2001 (first entry)XX
XX

XX *Corynebacterium glutamicum*; homeostasis; adaptation; HA protein;
KW fine chemical production; organic acid; proteinogenic amino acid;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
KW nucleotide; aromatic compound; vitamin; cofactor; polyketide; enzyme;
KW carboxylate; *Corynebacterium diphtheriae*; genetic engineering;
KW diagnosis; *Corynebacterium* diphtheriae; environmental condition;
KW *Corynebacterium*; environmental condition;

XX
OC
cornebacterium glutamicum.

XX
DN
W0200100842-A2:XX
PD
04-JAN-2001XX
DE 23 - JUN - 2000: 2000WO-1B00911.

XX 25 JUN 1999. 99US-0141031.

PR 08-JUL-1999; 99DE-1031636.
08-JUL-1999; 99DE-1032125.
08-JUL-1999; 99DE-1032125.

PR 09-JUL-1999; 99DE-1032126.

09-JUL-1999; 99DE-1032128.

(pIDD) that induces apoptosis. p53 can hold cells in check, quiescent and prevent cells from becoming cancerous. pIDD comprises leucine-rich repeat (LRR), antigenic fragments and p53-binding sequence which is responsive to p53. pIDD is useful for treating or preventing diseases associated with cell cycle regulation, cell growth and deficient or insufficient apoptosis, autoimmune disease, neurodegenerative disorder, dysplasia and Li-Fraumeni syndrome.

xx	Sequence	915 AA;
50		

Query Match	6.1%	Score 92.5;	DB 22;	Length 915;
Best Local Similarity	26.7%;	pred. No. 0.72;		
				Gaps 9;
				Indels 27;

[illegible]

RESULT 11

AAG90354

XX AAG90354:

XX 26-SEP-2001 (first entry)

XX
a glutathione protein fragment SEQ ID NO: 4108.

XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.

XX CG Corynebacterium glutamicum.

XX
DN
FP1108790-A2.

20-JUN-2001.

XX
DE 19-DEC-2000: 2000EP-0127688.

XX 16-DEC-1000. 99.TP-0377484.

PR 07-APR-2000; 2000JP-0159162.

XXXXXX

XXXXXX (K) XXXXXXXX

PI Nakagawa S, Mizoguchi M, Ikeda
PI Tateishi N, Senoh A, Ikeda

XX
 2001-276931/A0

DR N-PSDB; AAH65573.

Novel polynucleotides derived from Corynebacterium bacteria, for identifying

PT mutation point of a gene, measuring expression of a gene, analysing

PT expression profile or pattern of a gene and identifying homologous gene

XX	-	-	-	4708.	246nn + sequence	Listing; English.
XX	-	-	-			

XX The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium *Corynebacterium glutamicum*. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived

[illegible]

Tue Jan 15 14:12:19 2002

CC used for the recombinant production of epothilones in a heterologous host
CC that is more amenable to fermentation.

XX Sequence 1421 AA;

Query Match 5.9%; Score 90; DB 21; Length 1421;
Best Local Similarity 21.5%; Pred. No. 2.6;
Matches 78; Conservative 37; Mismatches 122; Indels 126; Gaps 19;

QY 5 VCLTVGPEVFRVTLSRYRALSPAEREAKLLDLY-----SSNIKTLRGAADYCA 54
DB 129 vfqigp-----seyeaalpqtasaeidagggigtmsvgagrisvalgrpcv 179
QY 55 AHDRLRYRLSSFLPMLDLAGDDTGAAYTHLAPQLLEAGHAFT--DAGVRLLMHPEQFI 112
DB 180 avdt---aysssl-----vavhlacqslrsgcestalagvslmlspstlv 222
QY 113 VLNSDRPEVRESSVRAMSAHARVMDGLGLARTPWMLLL--LHGK--GGR----- 158
DB 223 wlsktralardgrckafsaee---dgfrgegcavvvlkrlsgradgrlilavirgsai 279
QY 159 ---GAELAALIPD-----LPDPVRLR-----L 177
DB 280 nhdgassltvongssqeqivikraladagcaassvgvyeahgtgttlgdpieiqalnavy 339
QY 178 GLENDERAYSPAELLPICEATGTP-----LVFDAHHVHVHDKLPDQE--DPSV-- 223
DB 340 gjgrd--vatpllgsvktnighpeyasgitgllkvvslsqhgqipahlhaqalnprsw 397
QY 224 ---REWVLARATWQPPEWVHV-----LSNGIEGPQDRRHSHLI--ADFPsAYADVPOIEV 275
DB 398 gdlrlvttrtpw--pdwntpragvssfgmsg----tnahvvleapaaatctppaper 451
QY 276 EAK 278
DB 452 pae 454

RESULT 14
AAW52844
ID AAW52844 standard; Protein; 1891 AA.

XX AAW52844;

XX 24-JUL-1998 (first entry)

XX Amycolatopsis mediterranei strain wt3136 5.7 kb KpnI fragment protein.

XX Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
XX polyketide synthase; actinomycete; ansamycin.

XX Amycolatopsis mediterranei.

XX W09807868-A1.

XX 26-FEB-1998.

XX 18-AUG-1997; 97WO-EP04495.

XX 20-AUG-1996; 96EP-0810551.

XX (NOVS) NOVARTIS AG.

XX Engel N, Schupp T, Toupet C;

XX WPI; 1998-169172/15.

XX N-PSDB; AAV21186.

XX Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
XX to produce rifamycin and rifamycin analogues

XX Example 6; Page 43-53; 205pp; English.

XX The present sequence represents Amycolatopsis mediterranei strain wt3136
CC 5.7 kb KpnI fragment protein, from the present invention. The present
CC invention describes a Amycolatopsis mediterranei rifamycin synthesis
CC gene cluster DNA fragment comprising a DNA region involved directly or
CC indirectly in the gene cluster responsible for rifamycin synthesis,
CC including the adjacent DNA regions to the right and left which, by
CC reason of their function in connection with rifamycin biosynthesis,
CC qualify as constituents of this rifamycin gene cluster, and functional
CC fragments, derivatives or constituents of these. The Amycolatopsis
CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC for producing rifamycin, rifamycin analogues or precursors. It can also
CC be used for inactivating or modifying genes involved in ansamycin or
CC rifamycin biosynthesis. The DNA can be used for constructing mutant
CC actinomycetes strains from which the natural rifamycin or ansamycin
CC biosynthesis gene cluster has been partly or completely deleted. The
CC DNA fragment can be used for assembling a library of polyketide
CC synthases, which can be used for assembling a library of polyketides.
CC A hybridisation probe of the invention can be used for identifying DNA
CC fragments involved in the biosynthesis of ansamycins.

XX Sequence 1891 AA;

Query Match 5.9%; Score 90; DB 19; Length 1891;
Best Local Similarity 22.3%; Pred. No. 4.1;
Matches 54; Conservative 21; Mismatches 61; Indels 106; Gaps 10;

QY 22 RYRALSPAERAKLLDLYSSNIKTLRGAADYCAAHDIRLYRLSSLPMLDLAGDDTGA 81
DB 1023 rlaglapaeaeallldvrtqv-----al 1046
QY 82 VLTHLAPQLLEAGHAFTDAGVRLLMHPEQFTVLNSDRPEVRESSVRAMSAHARVMDGLGL 141
DB 1047 vlgahgepeavradtafktdgdsfslsvel-----rnlreas-----gkl 1087
QY 142 ARTPMNLLLLHGKGGKRGAEALALIPDLPVRLRLGLEN---DERAYSPAELLPICEAT 198
DB 1088 p-----atlvdytpvalarylrdefgdtvattptvataaada- 1126
QY 199 GTPLVFDAHVVHVDKLPDQEDPSVREWLRARATWQPPE---WQVHLSNGIEG----PQ 252
DB 1127 gepi-----aivgmactlpggvtqpegilwrlvr--dqleglsipie 1165
QY 253 DR 254
DB 1166 dr 1167

RESULT 15

AAW52849

ID AAW52849 standard; Protein; 3413 AA.

XX AAW52849;

XX 24-JUL-1998 (first entry)

XX A. mediterranei rifamycin synthesis gene cluster fragment protein E.
XX Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
XX polyketide synthase; actinomycete; ansamycin.

XX Amycolatopsis mediterranei.

XX W09807868-A1.

XX 26-FEB-1998.

XX 18-AUG-1997; 97WO-EP04495.

XX 20-AUG-1996; 96EP-0810551.

XX (NOVS) NOVARTIS AG.

XX Engel N, Schupp T, Toupet C;
 XX WPI; 1998-169172/15.
 DR N-PSDB; AAV21187.
 XX
 PT Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
 PT to produce rifamycin and rifamycin analogues
 XX
 PS Claim 6; Page 170-187; 205pp; English.
 XX
 CC The present sequence represents a Amycolatopsis mediterranei rifamycin
 CC synthesis gene cluster ORF E protein from the present invention. The
 CC DNA fragment comprises a DNA region involved directly or indirectly
 CC in the gene cluster responsible for rifamycin synthesis, including
 CC the adjacent DNA regions to the right and left which, by reason of
 CC their function in connection with rifamycin biosynthesis, qualify
 CC as constituents of this rifamycin gene cluster, and functional
 CC fragments, derivatives or constituents of these. The Amycolatopsis
 CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used
 CC for producing rifamycin, rifamycin analogues or precursors. It can also
 CC be used for inactivating or modifying genes involved in ansamycin or
 CC rifamycin biosynthesis. The DNA can be used for constructing mutant
 CC actinomycetes strains from which the natural rifamycin or ansamycin
 CC biosynthesis gene cluster has been partly or completely deleted. The
 CC DNA fragment can be used for assembling a library of polyketide
 CC synthases, which can be used for assembling a library of polyketides.
 CC A hybridisation probe of the invention can be used for identifying DNA
 CC fragments involved in the biosynthesis of ansamycins.
 XX
 SQ Sequence 3413 AA;

Query Match 5.9%; Score 90; DB 19; Length 3413;
 Best Local Similarity 22.3%; Pred. No. 9.9;
 Matches 54; Conservative 21; Mismatches 61; Indels 106; Gaps 10;

QY 22 RYRALSPAEREAKLLDYSSNIKTLRGAADYCAAHDIRLYRLSSSLFPMLDLAGDDTGAA 81
 Db 1597 riaglapaeqeallldvrtqv-----al 1620

QY 82 VLTHLAPOLLEAGHAFTDAGVRLMHPEQIVLNSDRPEVRESSVRAMSAHARVMDGLGL 141
 Db 1621 vlhagpeavradtafdtdgtsvel-----rnlreas-----gkl 1661

QY 142 ARTFWNLLLHGKGGRGAEIAALIPDLPPVRLRLGLEN---DERAYSPAELLPICAT 198
 Db 1662 p-----atlvdptvalarylrdelgdtvattptvataaada- 1700

QY 199 GTPLVFDAAHHVVDKLPDOEDPSVREWVLRARATWQPPE--WQVVHLSNGIEG----PQ 252
 Db 1701 gepi-----aivgmacrlpggvtdeglwrlvr--dglelspfp 1739

QY 253 DR 254
 Db 1740 dr 1741

Search completed: January 15, 2002, 13:50:18
 Job time: 307 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:50:57 ; Search time 57.72 Seconds
(without alignments)
233.532 Million cell updates/sec

Title: US-09-724-296-2_COPY_230_828
Perfect score: 3112
Sequence: 1 MRLLRNTQISKRVFTL.....KRLTARKRSRKEVEDEK 599

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pap.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pap.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pap.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pap.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pap.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	140.5	4.5	733	3	US-08-725-459B-23
2	134.5	4.3	753	3	US-08-725-459B-25
3	134.5	4.3	753	3	US-08-725-459B-26
4	133.5	4.3	1588	5	PCT-US93-07261-11
5	133.5	4.3	1663	5	PCT-US93-07261-16
6	132	4.2	414	1	US-07-667-276A-4
7	130	4.2	1104	4	US-08-923-992A-4
8	128.5	4.1	769	3	US-08-725-459B-37
9	128.5	4.1	769	3	US-08-725-459B-38
10	127.5	4.1	769	3	US-08-725-459B-35
11	127.5	4.1	769	3	US-08-725-459B-36
12	127.5	4.1	769	3	US-08-725-459B-39
13	127.5	4.1	1964	2	US-08-790-912-3
14	127.5	4.1	2052	2	US-08-790-912-2
15	127.5	4.1	651	1	US-08-769-309A-17
16	127	4.1	651	3	US-08-994-570-17
17	127	4.1	651	3	US-08-994-570-17
18	127	4.1	708	4	US-08-235-836C-76
19	127	4.1	1780	1	US-08-769-309A-5
20	127	4.1	1780	3	US-08-994-570-5
21	127	4.1	2308	1	US-08-015-973-1
22	127	4.1	2308	1	US-08-448-164-1
23	127	4.1	2308	4	US-08-081-929-2
24	125.5	4.0	680	3	US-08-725-459B-3
25	125.5	4.0	730	3	US-08-725-459B-2
26	125.5	4.0	733	3	US-08-725-459B-29
27	125.5	4.0	733	3	US-08-725-459B-30

28	125.5	4.0	773	1	US-08-524-757-6
29	125.5	4.0	773	3	US-08-725-459B-1
30	125.5	4.0	773	3	US-08-725-459B-79
31	124	4.0	1093	5	PCT-US93-03077-1
32	123.5	4.0	733	3	US-08-725-459B-22
33	123.5	4.0	770	2	US-08-209-521-13
34	123	4.0	770	4	US-08-961-810-123
35	123	4.0	770	4	US-08-352-902D-123
36	121	3.9	770	4	US-08-235-836C-66
37	121	3.9	748	3	US-08-725-459B-24
38	121	3.9	800	6	5183745-3
39	120	3.9	748	3	US-08-725-459B-27
40	119.5	3.8	694	1	US-08-339-152A-18
41	119.5	3.8	694	2	US-08-007-999B-5
42	119.5	3.8	694	2	US-08-689-276A-5
43	119.5	3.8	695	1	US-08-123-702-2
44	119.5	3.8	695	1	US-08-339-152A-30
45	119.5	3.8	695	2	US-08-104-165-1

ALIGNMENTS

RESULT 1
US-08-725-459B-23
; Sequence 23, Application US/08725459B
; Patent No. 6084068
; GENERAL INFORMATION:
; APPLICANT: CONAWAY, RONALD C.
; APPLICANT: CONAWAY, JOAN W.
; TITLE OF INVENTION: ELONGIN A AND C FUNCTIONAL DOMAINS
; NUMBER OF INVENTIONS: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SIDLEY & AUSTIN
; STREET: 717 N. HARWOOD, SUITE 3400
; CITY: DALLAS
; STATE: TX
; COUNTRY: US
; ZIP: 75201-6507
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08725,459B
; FILING DATE: 04-OCT-1996
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: HANSEN, EUGENIA S.
; REGISTRATION NUMBER: 31,966
; REFERENCE/DOCKET NUMBER: 11146/07501
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 214-981-3300
; TELEFAX: 214-981-3400
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 733 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..733
; OTHER INFORMATION: /note= "amino acids 1-480 and
; OTHER INFORMATION: 521-773 of Elongin A"
; US-08-725-459B-23

Query Match 4.5%; Score 140.5; DB 3; Length 733;
Best Local Similarity 20.1%; Pred. No. 0.00083;
Matches 82; Conservative 75; Mismatches 167; Indels 83; Gaps 15;

```

QY 73 NSFSAEPTVDLKKENETELANISGPHKKSTSTSTRKRRARSKKKATDSVSDKIDESVASY 132
Db 243 SSHKEKRPVDARGDEKSSVMGREGSKSHKASSKEESRLL-----SEDSAKEKLPSSVVKK 296
QY 133 DSSTHLRRSSRSKKPVN--YNSSESESEEOISKATKKVKKEEEVEEVEEVEDEKSLKNES 190
Db 297 EKD---REGNSLKKLSPALDVASDNHFKPKHKDEKIKSKDKNKQSVDSVD-----S 346
QY 191 SSDFEPV-----VPEOLETPISKKRRS--RSSAKNLEKESTMNLDHAPR-----EMF 237
Db 347 GRGTGDPDLPRAKDKVPNNLKAQEGKVRTNSDRKSPGLPKVEEMDMDEFEQPTMSFESY 406
QY 238 DCLDKPIPWGRGLGYACNLTLILSMKERVFCSTCRITTIORDGLESVKQLGT-QNVLDL 296
Db 407 LSYDQP-----RKKKKVVKTSG---TALGEKGLKKKSKSTSKNLNSA 447
QY 297 IKLVENHNFIHFH-----RVSSDLFPFA-----SHAKYGYTLEF 332
Db 448 QKLPKANENKSDKLQAGAEPTRPVKVPTDVLPEAGFTGRMNKMQVYSGSKCAYLPLKM 507
QY 333 AQSHLEEVGKLANKYNHRLTMHPGQVTOIASPREVVVDSAIRDLAYHDEILSRMKLNEOL 392
Db 508 MTUHQOCIRVLKNNIDISIEFVGVPYSVL-----EPVLERCTPDQLYRIEENHVLIEE-- 561
QY 393 NKDAVLIHLGGTF-EGKKEITLDRFRKNYORLSDSVKARLVLENDV 438
Db 562 -TDQLMKVCHRDPEERPEEYESWREMYLRLQDAREQRLRLTNNI 607

```

RESULT 2

```

US-08-725-459B-25
; Sequence 25, Application US/08725459B
; Patent No. 6084068
; GENERAL INFORMATION:
; APPLICANT: CONAWAY, RONALD C.
; APPLICANT: CONAWAY, JOAN W.
; TITLE OF INVENTION: ELONGIN A AND C FUNCTIONAL DOMAINS
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SIDLEY & AUSTIN
; STREET: 717 N. HARWOOD, SUITE 3400
; CITY: DALLAS
; STATE: TX
; COUNTRY: US
; ZIP: 75201-6507
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08725,459B
; FILING DATE: 04-OCT-1996
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: HANSEN, EUGENIA S.
; REGISTRATION NUMBER: 31,966
; REFERENCE/DOCKET NUMBER: 11146/07501
; TELEPHONE: 214-981-3300
; TELEFAX: 214-981-3400
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 753 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..753
; OTHER INFORMATION: /note="amino acids 1-545 and

```

```

; OTHER INFORMATION: 566-773 of Elongin A"
; US-08-725-459B-25

```

Query Match

```

Best Local Similarity 4.3%; Score 134.5; DB 3; Length 753;
Matches 86; Conservative 73; Mismatches 169; Indels 95; Gaps 18;

```

```

QY 73 NSFSAEPTVDLKKENETELANISGPHKKSTSTSTRKRRARSKKKATDSVSDKIDESVASY 132
Db 243 SSHKEKRPVDARGDEKSSVMGREGSKSHKASSKEESRLL-----SEDSAKEKLPSSVVKK 296
QY 133 DSSTHLRRSSRSKKPVN--YNSSESESEEOISKATKKVKKEEEVEEVEEVEDEKSLKNES 190
Db 297 EKD---REGNSLKKLSPALDVASDNHFKPKHKDEKIKSKDKNKQSVDSVD-----S 346
QY 191 SSDFEPV-----VPEOLETPISKKRRS--RSSAKNLEKESTMNLDHAPR-----EMF 237
Db 347 GRGTGDPDLPRAKDKVPNNLKAQEGKVRTNSDRKSPGLPKVEEMDMDEFEQPTMSFESY 406
QY 238 DCLDKPIPWGRGLGYACNLTLILSMKERVFCSTCRITTIORDGLESVKQLGT-QNVLDL 296
Db 407 LSYDQP-----RKKKKVVKTSG---TALGEKGLKKKSKSTSKNLNSA 447
QY 297 IKLVENHNFIHFH-----RVSSDLFPF-----ASHAKYGY--YTFEFAQS-- 335
Db 448 QKLPKANENKSDKLQAGAEPTRPVKVPTDVLPAIQTNYRPLPSLELISFQ 507
QY 336 -----HLEEVGKLANKYNHRLTMHPGQ---YTOIA-----SPREVVVDSAIRDL 376
Db 508 PKRAFSSPOEEEAAGTGRMNKMQVYSGSKCAYLPFEVGGVYSLPEVLERCTPDQ 567
QY 377 AYHDEILSRMKLNEOLNKDAVLIHLGGTF-EGKKEITLDRFRKNYORLSDSVKARLVLEN 435
Db 568 LYRIEENHVLIEE---TDQLMKVCHRDPEERPEEYESWREMYLRLQDAREQRLRLLT 624
QY 436 DDV 438
Db 625 NNI 627

```

RESULT 3

```

US-08-725-459B-26
; Sequence 26, Application US/08725459B
; Patent No. 6084068
; GENERAL INFORMATION:
; APPLICANT: CONAWAY, RONALD C.
; APPLICANT: CONAWAY, JOAN W.
; TITLE OF INVENTION: ELONGIN A AND C FUNCTIONAL DOMAINS
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SIDLEY & AUSTIN
; STREET: 717 N. HARWOOD, SUITE 3400
; CITY: DALLAS
; STATE: TX
; COUNTRY: US
; ZIP: 75201-6507
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08725,459B
; FILING DATE: 04-OCT-1996
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: HANSEN, EUGENIA S.
; REGISTRATION NUMBER: 31,966
; REFERENCE/DOCKET NUMBER: 11146/07501
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 214-981-3300
; TELEFAX: 214-981-3400

```


NAME: Blasdale, John H. C.
REGISTRATION NUMBER: 31,895
REFERENCE/DOCKET NUMBER: DX0288K
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-822-7398
TELEFAX: 201-822-7039
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1588 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-07261-11

Query Match 4.3%; Score 133.5; DB 5; Length 1588;
Best Local Similarity 19.4%; Pred. No. 0.011;
Matches 108; Conservative 84; Mismatches 175; Indels 191; Gaps 26;

QY 83 LKKNETELANISGPHKKSTSTSTRKARSKKKATDSVSKIDESVASYDSSSTHLRRSS 142
Db 7 LKQKTE-----KNEKARNALKEKLLKQKKDAOKAKDLTKKE--SQDSS-----SE 51
QY 143 RS-KKPNYNSSSESEEQI-SKATKKVKQKEEVEEVEDEKSLKNESSSDEFEPVVP 200
Db 52 KSLKEKVNGEALKEKENKETLKKKELENOKEKEBKNTKNNDEALANKGNDKDDKILVP 111
QY 201 EOLTPISKRRSSAKNLEKSTMNLDHAPMEFDCLDKPIPWGRGLGYACNLTLR 260
Db 112 KKPESVEKDLKEMELKEPEFIQ--HLKDYEEK--EKRNRW-----ILR 152
QY 261 SMKERVFCSTRTITTTORGLGESVKQLGTQNVLDLKLKVEWHNFGIHFMRVSSDLFPF 320
Db 153 SLR-----RDKLREIQ-----165
QY 321 ASHAKYGYTLFAOHLLEVCKLANKYNNHRLTMHPGOVTOIASPREVVVDSAIRDLAYHD 380
Db 166 ---EKLNAQLESAINELKE--RRASRRPMVMYKMOGMKDE-----VDEWK--KYDD 210
QY 381 EILSMKLNQNLKDAVLIIHLGGTFEGKKE-----LDREKKNYQRLSDSVK 428
Db 211 EQAEKNGTKDEIKDK-----GDQYEEIVETKFGYMRENALGELDEYEEVE-----K 258
QY 429 AKLVLENDVSVSVDLILPQELNIPVLVLDHNNHIVPGTLREGSLDLMLPIPIRETW 488
Db 259 KRYILK-EDGEGLDK-----VEEKLEETGYGFEKFPTRILV 296
QY 489 TRKGITQKHVSESADPTAISG-----MKRRAHSDRVF-----DFPPCDPTMDL 532
Db 297 KRRKKEQKLLKEDKEKKLIAAEEPPDEKKIKLSDSDKVVVVPVNNKSSFPD-----349
QY 533 MIEAKEQOAVFELCRRYELQN--PPCP-----LETMGPEYDQTRDGYYPGAEKRLTA- 584
Db 350 KTRAPDKKTMF-----YRLSELFIVPRKONELAVCGDSMDSKVNG-----KKLAKST 397
QY 585 -----KRRRSKKEEVEHDE 598
Db 398 FNPFRNRNKLKERKMOE 415

RESULT 5
PCT-US93-07261-16
Sequence 16, Application PC/TUS9307261
GENERAL INFORMATION:
TITLE OF INVENTION: PEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: John H. C. Blasdale
CITY: Madison
STATE: New Jersey
COUNTRY: USA
ZIP: 07940-1000

INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 753 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Protein
LOCATION: 1..753
OTHER INFORMATION: /note="amino acids 1-565 and
OTHER INFORMATION: 586-773 of Elongin A"
US-08-725-459B-26

Query Match 4.3%; Score 134.5; DB 3; Length 753;
Best Local Similarity 19.8%; Pred. No. 0.0029;
Matches 83; Conservative 75; Mismatches 173; Indels 89; Gaps 16;

QY 73 NFSFAETPVDLKAKENETELANISGPHKKSTSTSTRKARSKKKATDSVSKIDESVASY 132
Db 243 SSHKEKRPVDARGDEKSSVMGREKSHKASSKEESRLL-----SEDSAKEKLPSVVYKK 296
QY 133 DSSTHLRRSSRSKKPN--YNSSESESEEQISKATKKVKQKEEVEEVEDEKSLKNES 190
Db 297 EXD---REGNSLKKLSPALDVASDNHFKKPKKDSEKIKSKDNKQSVDSVD-----S 346
QY 191 SDEFPVP-----VPEQLETPISKRRRS--RSSAKNLEKSTMNLDHAPR-----EMF 237
Db 347 GRGTGDLPRAKDKVNNLKAQEGVYRTNSDRKSPGSLPKVEEMDMDFEQPTMSFESY 406
QY 238 DCLDKPIPWGRGLGYACNLTLRSMKRVFCSTRTITTTORGLGESVKQLGT-ONVLDL 296
Db 407 LSYDQP-----RKKKKVVKTSG---TALGEKGLKKDKSKSTKNLSA 447
QY 297 IKLVENHNHFGIHF-----RVSSDLFPF-----ASHAKYG--YTLEFAQS-- 335
Db 448 QKLPKANENKSDKLQAPAGAEPTPRKVPDVLPAIPALQTNRYRPLPSLELISSFQ 507
QY 336 -----HLEVEGKLANKYNNHRLTMHPGOVTOIASPREVVVDSAIRDLAYHDEILSR 385
Db 508 PKRKAFSSPOEEEGEAGFTGRMNSKQVYSGSKCAYLPKMMTLHQQCIRVLKNNIDSIDQ 567
QY 386 MKNLEQNL-----KDAVLIIHLGGTF-EGKKEITLDRFRKNYORLSDSVKARLVLENDV 438
Db 568 LYRIECNHVLIETDLQKVVHCHDFKEERPEYESWREMYLRLQDAREQLRLTLNMI 627

RESULT 4
PCT-US93-07261-11
Sequence 11, Application PC/TUS9307261
GENERAL INFORMATION:
TITLE OF INVENTION: PEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: John H. C. Blasdale
CITY: Madison
STATE: New Jersey
COUNTRY: USA
ZIP: 07940-1000
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 6.0.5
SOFTWARE: Microsoft Word 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07261
FILING DATE: 19930805
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/927,531
FILING DATE: 07-AUG-1992
ATTORNEY/AGENT INFORMATION:

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 6.0.5
SOFTWARE: Microsoft Word 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07261
FILING DATE: 19930805
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/927,531
FILING DATE: 07-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Blasdale, John H. C.
REGISTRATION NUMBER: 31,895
REFERENCE/DOCKET NUMBER: DX0288K
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-822-7398
TELEFAX: 201-822-7039
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1663 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
STRAIN: Malayan Camp
PCT-US93-07261-16

```

Query Match      4.3%; Score 133.5; DB 5; Length 1663;
Best local Similarity 19.4%; Pred. No. 0.012;
Matches 108; Conservative 84; Mismatches 175; Indels 191; Gaps 26;

QY      83 LKKNETELANISGPHKSTSTSRKRARGSSKKATDSVDKIDESVASYDSSPHLRSS 142
Db      7 LKQKTE-----KNEARNALKEKLLKQKNDQAOKADLTKKE---SQDSS-----SE 51

QY      143 RS-KKPNYNSSSSESEEQI-SKATKKVYQKEEEVVEEDVDEKSLKNSSSDBEFPVPV 200
Db      52 KSLKEKVGNGALKKEKENETLKKLELNEQKEKEEKNKIKNDNDALKNGKNDKDKKIVP 111

QY      201 EQETPIKRRRRKSSAKNLEKESTMNLDDHAPREMFCDLCKPIPWGRGLGYACLTILR 260
Db      112 KPESVEKDLKEMELKEKETIKQ---HLKDYEBEK-----EKRRNW-----ILR 152

QY      261 SMKERVCSRTCRTITQRDGLSVKOLGTQNVLDLIKLVENHNFGIHFMRVSSDLPFF 320
Db      153 SLR-----RKLREIQEL-----165

QY      321 ASIAKYGYTLFEQAOSLLEEVGKLANKYNHRLTMHPGOYTQIASPREVVVDSAIRLAYHD 380
Db      166 ---EKLNAQESAINELKE--RRASRPMVMVKMORGKDE-----VDEWIK--KYDD 210

QY      381 EILSRMKNLQNLKANVLIIHLGTFEGKET-----LDPRKNYORLSDSVK 428
Db      211 EQAEKNGTKDEEKDK-----GDGYEIVETFKYGMRENALGEDEYERIE-----K 258

QY      429 ARVLVENDVSWSVQDILLPCQELNIPVLVDWHHNIVPGTLREGSLDMLPTIETW 488
Db      259 KRYLK-EDGEGDLKD-----VEEKLEETGYCFREKFPPTIRLV 296

QY      489 TRKGITQKHYSADPTAISG-----MKRRASHDRVF-----DFPCCDPTMDL 532
Db      297 KRRNRKEQKLLKEDKEKLLAAEEPDDKEKKILKDSDDKVVVPVNNKSSFPD-----349

QY      533 MIEAKEKEQAVELCRRYELQN--PPCP-----LEIMGPEYDQTRDGYPPGAERKRLTA- 584
Db      350 KFPADKKRTMF-----YRLSELFPIVPRKDNELAVCGDSMDSKVNG-----KKLKST 397

QY      585 ----KRRRSRKEEVEEDE 598
Db      398 FNPFKRRNRNKLKERKMOE 415

```

[illegible]

ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/923,992A
FILING DATE: 05-SEP-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/024,707
FILING DATE: 06-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1104 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-923-992A-4

Query Match 4.2%; Score 130; DB 4; Length 1104;
Best Local Similarity 19.7%; Pred. No. 0.013;
Matches 119; Conservative 88; Mismatches 187; Indels 210; Gaps 29;
QY 44 HCLPTLSLPLMSKTTLSMLPQVINGANSFSAETPVDLKKENETELANISGPHKKS-- 101
DB 142 HNKPNQOKTLKOSDFKVDLS-----NIDKELNHQKSOV---EAMAEQAGTNEDKRDSML 192
QY 102 -TSTTRKRRSSKKKATDSYDKIDSVASYDSSTHLRRSSRSKPKVYNSSSESES-- 158
DB 193 KIEDIRKQQAQADKKEAEV--KYRELKGLFSTKAGLDQOIOEHVKKETSSEENTQK 250
QY 159 -EQISKATKKYKQKEE-----EYVEEVDKSLKNSSSDEPEPVPEQLETPISK 209
DB 251 VDEHYANSIQNLAQKSLSELDKATTNEQATQVKNQFLENQAKLKEIOPLIK---ETNVKL 307
QY 210 RRRSSRSKLNKE-----STMNLDHAPREMFDCLDKPI--PWGRGLGYACLNILRSKM 263
DB 308 YKAMESLEQVEKELKHNEANLQDLVAK-----SKELVREYEGKLN----- 349
QY 264 ERVFCSTCRITTIQDGLGIESVQKLTQNVLDLIKLEWNNHFGIFHFMVSSDLPFPASH 323
DB 350 -----QSKNLPKQLQEEEAHSLKQVVE-----HFRK----- 377
QY 324 AKYGYTLFAQSHLEEVGKLANKYNHRLTMHPGQVQTATSPREVVDVSAIBOLAYHDEIL 383
DB 378 -----KFKTS--EQV-----TPKKRVK-----RDLAANEN-- 400
QY 384 SRMKLNQLKNAQVLIHLGGTFEGK-----KETL-----DRFR 417
DB 401 NQKIETLVSPENITV-----YEGEDVKFTYAKSDSKTTLDFSLTLTKYNPSVDRLS 454
QY 418 KNYQLRSDSYK--ARLVLENDVDSWSVODLLPLCOELNIPLDVLDHNNHIVPG-----TLRE 472
DB 455 TNYKTNTDNHIAEITKN-----LKNQSQTVTLKAKDSDGNVVEKTFITVOK 504
QY 473 GSLDMLPIPTIRETWTRKGIQKHYSSEADPTAISGKMRRAHSDRVDFPFCPTMDL 532
DB 505 KEERQVPTPTPEQKHSKTEQNVPOE---PKSNQKNQLOELIKSAQO----- 546

QY 533 MIEAKEKEQAVFELCRRRYELQNPQCPLEIMGPYDQTRDGYPPGAERKRLTARKRRSRKE 592
DB 547 --ELEKLEKAIKEL-----MEQPEIP---SNPEY-----GIQKSLW-----ESOKE 582
QY 593 EVEE 596
DB 583 PQE 586

RESULT 8

US-08-725-459B-37
Sequence 37, Application US/08725459B
Patent No. 6084068
GENERAL INFORMATION:
APPLICANT: CONAWAY, RONALD C.
APPLICANT: CONAWAY, JOAN W.
TITLE OF INVENTION: ELONGIN A AND C FUNCTIONAL DOMAINS
NUMBER OF SEQUENCES: 79
CORRESPONDENCE ADDRESS:
ADDRESSEE: SIDLEY & AUSTIN
STREET: 717 N. HARWOOD, SUITE 3400
CITY: DALLAS
STATE: TX
COUNTRY: US
ZIP: 75201-6507
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/725,459B
FILING DATE: 04-OCT-1996
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: HANSEN, EUGENIA S.
REGISTRATION NUMBER: 31,966
REFERENCE/DOCKET NUMBER: 11146/07501
TELECOMMUNICATION INFORMATION:
TELEPHONE: 214-981-3300
TELEFAX: 214-981-3400
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 769 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: protein
LOCATION: 1..769
OTHER INFORMATION: /note= "amino acids 1-552 and
OTHER INFORMATION: 557-773 of Elongin A"
US-08-725-459B-37

Query Match 4.1%; Score 128.5; DB 3; Length 769;
Best Local Similarity 20.0%; Pred. No. 0.01;
Matches 88; Conservative 72; Mismatches 168; Indels 111; Gaps 18;
QY 73 NPSAETPVDLKKENETELANISGPHKKTSTSTRKRARSSKKKATDSVSKIDESVASY 132
DB 243 SSHKRRPVDARDEKSSVWGREGKSHKASSKEESRLL-----SEDSAKEKLPSSVVK 296
QY 133 DSSTHLRRSSRSKPKVYN--YNSSSSESEEQISKATKKVKQKEEEYVEEVDKSLKNES 190
DB 297 EKD--REGNSLKKLSPALDVAQDNHFKPKHDKSEKIKSDKNQKQSDVD-----S 346
QY 191 SSDEFEPV-----VPEQLFTPISEKRRS--RSSAKNLEKSTMNLDHAPR-----EMF 237
DB 347 GRUTGDLPLRAKDKVNPNNLKAQEGKVRTNSDRKSPGSLPKVEEMDMDDFEQPTMSFESY 406

```

QY 238 DCLDKPIPMRGRGLGYACLTILRSMKERVFCSTRTITTIORDGLSVKOLGT-QNVLDL 296
Db 407 LSYDQP-----RKKKKVVKTSG---TALGEKGLKKKDSKSTKNLSA 447
QY 297 IKLVENHNFIHFM-----RVSSDLFPF-----ASHAKYG--YTFLEFAQS-- 335
Db 448 QKLPRANENKSDKLQAPAGAEPTPRKVPDVLPAIPALPOTNYRPLPSLELISSFO 507
QY 336 -----HLEVGKLANKYNHR-----LTMHPGQYTOIASPREV--- 367
Db 508 PKRAFFSPQEEBAGFTGRRMNSKMOVYSGSKAYLPKMMTLHQVKNNDISIFEVGGV 567
QY 368 -----VVDSALRDLAYHDEILSRMKLNQKDAVLIHILGGTF-EGKKTETLDRFRKN 419
Db 568 PYSVLEPVLERTCPDQLYRIEENHVLIEE---TDQLWKVCHCRDFKEERPEYESWREM 624
QY 420 YORLSDSVKARLVLENDV 438
Db 625 YLRLODAREQRLRLTNNI 643

RESULT 9
US-08-725-459B-38
; Sequence 38, Application US/08725459B
; Patent No. 6084068
; GENERAL INFORMATION:
; APPLICANT: CONAWAY, RONALD C.
; APPLICANT: CONAWAY, JOAN W.
; TITLE OF INVENTION: ELONGIN A AND C FUNCTIONAL DOMAINS
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SIDLEY & AUSTIN
; STREET: 717 N. HARWOOD, SUITE 3400
; CITY: DALLAS
; STATE: TX
; COUNTRY: US
; ZIP: 75201-6507
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/725,459B
; FILING DATE: 04-OCT-1996
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: HANSEN, EUGENIA S.
; REGISTRATION NUMBER: 31,966
; REFERENCE/DOCKET NUMBER: 11146/07501
; TELEPHONE: 214-981-3300
; TELEFAX: 214-981-3400
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 769 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: protein
; LOCATION: 1..769
; OTHER INFORMATION: /note="amino acid 1-556 and
; OTHER INFORMATION: 561-773 of Elongin A"
US-08-725-459B-38

```

```

Query Match 4.1%; Score 128.5; DB 3; Length 769;
Best Local Similarity 20.0%; Pred. No. 0.01;
Matches 88; Conservative 72; Mismatches 168; Indels 111; Gaps 18;
QY 73 NSFSAETPVDLKKENETELANISCPHKKSTSTSTRKRARRSKKATDSVSKIDESVASY 132

```

```

Db 243 SSHKEKRPVDARGDEKSSVMGRKSHKASSKEBSRLL-----SEDSAKEKLPSVYVK 296
QY 133 DSSTHLRRSSRSKKPVN--YNSSESESEEQISKATKKYKQKEEYVEEVEDEKSLKNES 190
Db 297 EKD---REGNSLRKKLSPALDVAENHFKPKHKDSEKIKSDKNKQSVDSVD-----S 346
QY 191 SSDEFEPV-----VPEQLETPISKRRRS--RSSAKNIKESTMNLDHAPR-----EMF 237
Db 347 GRGTGDPPLPRAKOKVFNLLKAQEGKVRTNSDRKSPGLPKVEEMDDDEFTQPMFSFESY 406
QY 238 DCLDKPIPMRGRGLGYACLTILRSMKERVFCSTRTITTIORDGLSVKOLGT-QNVLDL 296
Db 407 LSYDQP-----RKKKKVVKTSG---TALGEKGLKKKDSKSTKNLSA 447
QY 297 IKLVENHNFIHFM-----RVSSDLFPF-----ASHAKYG--YTFLEFAQS-- 335
Db 448 QKLPRANENKSDKLQAPAGAEPTPRKVPDVLPAIPALPOTNYRPLPSLELISSFO 507
QY 336 -----HLEVGKLANKYNHR-----LTMHPGQYTOIASPREV--- 367
Db 508 PKRAFFSPQEEBAGFTGRRMNSKMOVYSGSKAYLPKMMTLHQVKNNDISIFEVGGV 567
QY 368 -----VVDSALRDLAYHDEILSRMKLNQKDAVLIHILGGTF-EGKKTETLDRFRKN 419
Db 568 PYSVLEPVLERTCPDQLYRIEENHVLIEE---TDQLWKVCHCRDFKEERPEYESWREM 624
QY 420 YORLSDSVKARLVLENDV 438
Db 625 YLRLODAREQRLRLTNNI 643

RESULT 10
US-08-725-459B-35
; Sequence 35, Application US/08725459B
; Patent No. 6084068
; GENERAL INFORMATION:
; APPLICANT: CONAWAY, RONALD C.
; APPLICANT: CONAWAY, JOAN W.
; TITLE OF INVENTION: ELONGIN A AND C FUNCTIONAL DOMAINS
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SIDLEY & AUSTIN
; STREET: 717 N. HARWOOD, SUITE 3400
; CITY: DALLAS
; STATE: TX
; COUNTRY: US
; ZIP: 75201-6507
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/725,459B
; FILING DATE: 04-OCT-1996
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: HANSEN, EUGENIA S.
; REGISTRATION NUMBER: 31,966
; REFERENCE/DOCKET NUMBER: 11146/07501
; TELEPHONE: 214-981-3300
; TELEFAX: 214-981-3400
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 769 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Protein

```


RESULT 13
US-08-725-459B-40
; Sequence 40, Application US/08725459B
; Patent No. 6084068
; GENERAL INFORMATION:
; APPLICANT: CONAWAY, RONALD C.
; APPLICANT: CONAWAY, JOAN W.
; TITLE OF INVENTION: ELONGIN A AND C FUNCTIONAL DOMAINS

[illegible]

Tue Jan 15 14:12:08 2002

Db 625 YLRQDAREQRLRLTNNI 643

RESULT 14

US-08-790-912-3
; Sequence 3, Application US/08790912
; Patent No. 5976542
; GENERAL INFORMATION:
; APPLICANT: Weiser, Jeffrey N.
; APPLICANT: Plaut, Andrew G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
; TITLE OF INVENTION: OF STREPTOCOCCUS PNEUMONIAE INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: 1601 Market Street, 36th Floor
; City: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103-2398
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/790,912
; FILING DATE: 29-JAN-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,939
; FILING DATE: 23-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary, Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 7600-401
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 567-2020
; TELEFAX: (215) 567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1964 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-790-912-3

Query Match 4.1%; Score 127.5; DB 2; Length 1964;

Best Local Similarity 18.9%; Pred. No. 0.053;
Matches 86; Conservative 79; Mismatches 180; Indels 109; Gaps 18;

Qy 47 PDTLKSLPMSKTTLSMLPQVNIAGNSFSAETPVDLAKK-----NETELANIS-----G 96
Db 459 PDIVS--DKGEFQVAPLPEYKGNTEQVKPETPVEKTEQGPKEETVPVKPTETPVN 516
Qy 97 PHKKTSTSTRKARRSKKATDSV--SDKIDSVASYDSSTHLRRSSRSKPKVNYNSS 154
Db 517 PNEGTTEGTISQEAENPVQPAEESTTNSEKVSPTDTSSENTGE-----VSSN 562
Qy 155 ESEEEQISKATK----KVQKEEEYVEEV----DEKSLKNFSSSDSEFEPVVP-EQLET 205
Db 563 PSDSTTSVGSNKPENHDSKNSEKTEVEVPVNPNEGTVGTSNQETKPVQPAETQT 622
Qy 206 PISK-RRRSRSAAKNKESTMTLDHAPREMFCDLCKDPIWRGRGLYACLTILRSME 264
Db 623 NSGKIANENTGEVSNKPSDSKPPVEESNQEKNGTATKP-----HNFCHFM 311
Qy 265 RVFCSTRCTRIITQDGLGVSKOLGTONVLDLKLKLVN-----HNFCHFM 311
Db 662 -----ENSGNTTSNGQTEPEKKLELRNVDIELYSQTNGTYRQHVSLDGPENTDTYEV 716

Qy 312 RVSSDLF-----PFAS-----HAKYGYTLFEAQSHLEEVCKLANKYNHRLTMHPGQVQ 360
Db 717 KVKSSAFKDYIPIVASITEEKRNGQSVYKITAQAKLQO--ELENKYVDNFSFY----- 768
Qy 361 IASPREVVVDSAJRDLAYIDEILSRMKLNEQLNKDAVLIHLGGTTEGKKTLDREKRY 420
Db 769 -----LDKKAKE--ENTNFTSFSNLVKAQNQNSGYTHLAASLANANEVELGPDERSY 818
Qy 421 QRLSDSVKARLVLENDVSVSDLD-LPICQELN 453
Db 819 --IKDFTGRLIGEKDGKIYAIYNLKKPLFENLS 850

RESULT 15

US-08-790-912-2
; Sequence 2, Application US/08790912
; Patent No. 5976542
; GENERAL INFORMATION:
; APPLICANT: Weiser, Jeffrey N.
; APPLICANT: Plaut, Andrew G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
; TITLE OF INVENTION: OF STREPTOCOCCUS PNEUMONIAE INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: 1601 Market Street, 36th Floor
; City: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103-2398
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/790,912
; FILING DATE: 29-JAN-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,939
; FILING DATE: 23-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary, Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 7600-401
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 567-2020
; TELEFAX: (215) 567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2052 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-790-912-2

Query Match 4.1%; Score 127.5; DB 2; Length 2052;
Best Local Similarity 18.9%; Pred. No. 0.057;
Matches 86; Conservative 79; Mismatches 180; Indels 109; Gaps 18;

Qy 47 PDTLKSLPMSKTTLSMLPQVNIAGNSFSAETPVDLAKK-----NETELANIS-----G 96
Db 532 PDIVS--DKGEFQVAPLPEYKGNTEQVKPETPVEKTEQGPKEETVPVKPTETPVN 589
Qy 97 PHKKTSTSTRKARRSKKATDSV--SDKIDSVASYDSSTHLRRSSRSKPKVNYNSS 154
Db 590 PNEGTTEGTISQEAENPVQPAEESTTNSEKVSPTDTSSENTGE-----VSSN 635
Qy 155 ESEEEQISKATK----KVQKEEEYVEEV----DEKSLKNFSSSDSEFEPVVP-EQLET 205

```
Db 636 PSDSTTSVGSNKPEHNDKNSKENSEKTVEEVVNPNEGTVEGTSNQETKFPVQPAETQT 695
Qy 206 PISK-RRRSRSKWLKSTWNLDDHAPRPMFDCLDKPIWRCRLGYACLNTILRSMKE 264
Db 696 NSGKTANENTGEVSNKPSDSKPPVEESNQPEKNGTATKP----- 734
Qy 265 RVFCSRTCRTTIORDGLESVKQLGTQNVLDLIKLVEMN-----HNFQIHF 311
Db 735 -----ENSGNTTSENGQTEPEKKLELRNVSDIELYSQTNCTYRQHVSLDGIPTDTYFV 789
Qy 312 RVSSDLF-----PFAS-----HAKYGYTLEFAQSHLEEVGKLANKYNHRLTMHPCQYQ 360
Db 790 KVSSAFKDVYIPVASITEEKRNGQSVYKITAKAEKLOQ--ELENKYVDNFSFY----- 841
Qy 361 IASPREVVVDSAIRDLAYHDEILSRMKNLQNKDQAVLIHILGGTFEGKKEITLDRFRK 420
Db 842 -----LDKAKE--ENTNFTSFSNLVRAINONPSGTYHLAASLNANEVELGPDERSY 891
Qy 421 QRLSDSVKARLVLENDVSVSWODL-LPLQOELN 453
Db 892 --IKDTETGRLIGKDGKIYAIYNLKKPLFENLS 923
```

Search completed: January 15, 2002, 13:51:01
Job time: 330 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 14:07:00 ; Search time 45.07 seconds
(without alignments)
487.292 Million cell updates/sec

Title: US-09-724-296-2_COPY_230_828

Perfect score: 3112

Sequence: 1 MLRLKRNITQISKRIVFTIL.....KRLTARKRRSRKEVEDEK 599

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	360	11.6	320	1 YWJD_BACSU	P45864 bacillus su
2	160	5.1	2230	1 GOG4_HUMAN	Q13439 homo sapien
3	154	4.9	500	1 GAR2_SCHPO	P41891 schizosacch
4	150.5	4.8	726	1 YB1D_SCHPO	P87178 schizosacch
5	147	4.7	568	1 RRPP_PIIHE	P32532 human parai
6	145.5	4.7	918	1 IF38_SCHPO	O14164 schizosacch
7	145	4.7	607	1 YMH1_YEAST	Q04632 saccharomyc
8	144	4.6	1875	1 MLPI_YEAST	Q02455 saccharomyc
9	142.5	4.6	1462	1 NKCR_HUMAN	P30414 homo sapien
10	138	4.4	568	1 RRPP_PIIHD	P32531 human parai
11	138	4.4	1940	1 MYH3_HUMAN	P11055 homo sapien
12	136.5	4.4	1630	1 MSP1_PLAFK	P04932 plasmodium
13	136.5	4.4	1639	1 MSP1_PLAFW	P04933 plasmodium
14	135	4.3	1164	1 KEL1_YEAST	P38853 saccharomyc
15	134	4.3	2105	1 POLR_ASGVP	P36309 apple stem
16	133.5	4.3	1336	1 SEC3_YEAST	P33332 saccharomyc
17	132.5	4.3	1332	1 SPT7_YEAST	P35177 saccharomyc
18	132	4.2	414	1 NSR1_YEAST	P27476 saccharomyc
19	132	4.2	1790	1 USOL_YEAST	P25386 saccharomyc
20	131.5	4.2	952	1 YB15_CAEEL	P46012 caenorhabdi
21	130	4.2	1257	1 RKBI_HUMAN	P29374 homo sapien
22	129.5	4.2	1969	1 MYSA_CAEEL	P12844 caenorhabdi
23	129	4.1	1938	1 MYHD_HUMAN	O9ukx3 homo sapien
24	128.5	4.1	997	1 BIRL_SCHPO	Q14064 schizosacch
25	128	4.1	783	1 YAYB_SCHPO	Q10218 schizosacch
26	128	4.1	1781	1 KACB_HUMAN	Q02952 homo sapien
27	127.5	4.1	651	1 SEC9_YEAST	P40357 saccharomyc
28	127.5	4.1	845	1 NFM_BAT	P12839 rattus norv
29	127.5	4.1	1727	1 ALM1_SCHPO	O9uk5 schizosacch
30	127.5	4.1	2633	1 CENE_HUMAN	Q02224 homo sapien
31	127	4.1	663	1 MLH_TETTH	P40631 tetrahymena
32	127	4.1	747	1 DIL2_HUMAN	Q9ulw0 homo sapien
33	127	4.1	1744	1 TANA_XENLA	Q01550 xenopus lae

34	126.5	4.1	1420	1 APX_XENLA	Q01613 xenopus lae
35	126.5	4.1	1462	1 TOP2_PEA	Q24308 pisum sativ
36	126	4.0	450	1 ADPP_BOVIN	Q9tum6 bos taurus
37	126	4.0	997	1 SCPL_RAT	Q03410 raitus norv
38	125.5	4.0	964	1 XQY1_CAEEL	Q09560 caenorhabdi
39	125.5	4.0	4036	1 RRPL_DUGBV	Q66431 dugbe virus
40	125	4.0	230	1 YP23_CAEEL	Q09202 caenorhabdi
41	124.5	4.0	663	1 PIT_DROME	Q9vd51 drosophila
42	124.5	4.0	1704	1 VIT1_FUNHE	Q90508 fundulus he
43	124.5	4.0	2017	1 MYSN_DROME	Q95323 drosophila
44	124	4.0	1093	1 TMFL_HUMAN	P82094 homo sapien
45	124	4.0	2314	1 PTP2_HUMAN	P23471 homo sapien

ALIGNMENTS

```

RESULT 1
YWJD_BACSU
ID YWJD_BACSU STANDARD: PRT; 320 AA.
AC P45864;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 36.9 KDA PROTEIN IN ACDA 5'REGION.
GN YWJD.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Glaser P., de la Fuente V., Banchin A.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 249782; CAA89865.1; -
DR EMBL; 299123; CAB15748.1; -
DR Subtilist; BG11309; YWJD.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 320 AA; 36896 MW; 2FD35A710648E002 CRC64;

```

```

Query Match 11.6%; Score 360; DB 1; Length 320;
Best Local Similarity 29.8%; Pred. No. 9e-13;
Matches 84; Conservative 61; Mismatches 109; Indels 28; Gaps 8;
Qy 276 TIORDGLESVKGQNGQNVLDLIKLVENHNFHGFHFRVSSDLFPFASH--AKYGYTLFPA 333
Db 34 TERKEALLTVTKANLURNIM---RTLHYIIGHGIPLYRFSSSIVPLATHFDVWDFVTPF- 89
Qy 334 QSHLEEVGKLANKYNHRLTMHPGOYTQIASPREVVVDSAIRDLAYHDEILSRMKLINEQUN 393
Db 90 QKEPREICELVKTHGLRTSFHPNQFTLTSPKESVTKNVTDMAHYHMLEANGIADR-- 147
Qy 394 KDVALIIHLGTFEGKKTLDLRFKNYORLSVSKARLVLENDVDSWVSQDILLPLCQELN 453
Db 148 --SVNIHIGAYGNKDTATATQFQHNINIKQLPQEIHKRMTELENDKDTYTTTEETLQCEQED 205
Qy 454 IPLVLDDWHHNTVPCTLRREGSLDMLPLIPTIRETWKGTGKQHYSESADPTAISGMKR 513
Db 206 VPFVDFDHHFYANP-----DDHADLNVALPRMKTWERIGLQPKVHLSSPKSEQAI----- 256
Qy 514 RAHSDRV---FDPPPCD-----PTMDLMIEAKEKEQAVFEL 546

```


QY 494 TQKHYESADPTAISGMKRRASDRVFD 522
 DB 619 AQDIYWDETAKGTPI-----IERDVTVD 642

RESULT 5
 ID RPP_PILHE STANDARD; PRT; 568 AA.
 AC P32532;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE RNA POLYMERASE ALPHA SUBUNIT (EC 2.7.7.48) (NUCLEOCAPSID
 DE PHOSPHOPROTEIN).
 GN P.
 OS Human parainfluenza 1 virus (strain CI-14/83).
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
 OC Paramyxoviridae; Paramyxovirinae; Paramyxovirus.
 OX NCBI_TaxID=31606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92295573; PubMed=1318610;
 RA Power U.F., Ryan K.W., Portner A.;
 RT "The P genes of human parainfluenza virus type 1 clinical isolates
 RT are polyclonal and microheterogeneous.";
 RL Virology 189:340-343(1992).
 CC -!- FUNCTION: THIS PROTEIN IS PROBABLY A COMPONENT OF THE ACTIVE
 CC POLYMERASE. IT MAY FUNCTION IN TEMPLATE BINDING.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M74080; AAA46830.1; -;
 DR PIR; B40234; RNNZ83.
 DR InterPro: IPR002693; Paramyxo-P.
 DR Pfam; PF01806; Paramyxo-P; 1.
 KW Transferase; RNA-directed RNA polymerase; Nucleocapsid;
 KW Phosphorylation.
 SQ SEQUENCE 568 AA; 64586 MW; 8795D10C1CD3CCD7 CRC64;

Query Match 4.7%; Score 147; DB 1; Length 568;
 Best Local Similarity 21.2%; Pred. No. 0.41; Indels 122; Gaps 23;
 Matches 114; Conservative 87; Mismatches 214;

QY 62 LSLPQVNIANSFSAETPVDLKKENETELANISGPHKSTSTSTRKARRSKKATDSV 121
 DB 36 LSYKP-TEIGEDRWLHNI IDNPENKSKSDDDNNKDRALSTQDH-RSSEE---SGI 90
 QY 122 SDKIDESVSDSTHLR-----RSSR---SKKPVNNSSE-----SESEE 160
 DB 91 SRRTGES---KTETHARLDQOQIHASRGTSPPNLPENMGDERNTRIDEPSNRRH 146
 QY 161 QISKATKKVKO-----KEEYVEEYDEKSLKNSSSEDEFEPVVPQETPISKRRSR 214
 DB 147 QRSVPTDEDRKWAENSKREEDQVEGPEVRGTSLSDDGE-----GRIN 192
 QY 215 SSANKLEKSTMNLDIAPREMFCLDKPIPRWGRIGYACINTILSKMKRVCSTCRI 274
 DB 193 NNGRSMETSPT---HSTR-ITDVITNPS-----ELEPAVLQRKKRP 231
 QY 275 TTIORGLGESKOLGTONVLDLKLVEW-NHNFQIHFMRVSSDLFFPASHAKYGYLEFA 333
 DB 232 TTKRNOTRSE---TON-SELHKSTSGDSSNLENDHNTKTSQKIPP----- 273
 QY 334 QSHLEEYCKLANKYNNHRLTWHPGQYTOIASPREVYVDSATRDLAYHDEILSRMKLNEQLN 393
 DB 274 SKNEEPAVTKNNHNRKTKHT---TNNANNNAKCLPTPEHDTTSNEEGTSNTSDVEM-- 328

QY 394 KOAVLIHLG-----GTFEKKETLDRFRK-----NYQLSDSYKARLVLENDVSVS 442
 DB 329 --AKLLVSLGVNKSQHEPELSRRASHQFAKRLKSAKYKEMTNLCGMLLSVEKSLGNKV 386
 QY 443 QDLLPLCQELNIPVLVDWHHHHNIIVPTLREGSLDLMPLIPTIRETWTRKGTOKOHYES 502
 DB 387 EENRTLLKQIQEIDSSRLDKHFRSEYQKEQNSLMANLSTLHIITDRGKT-----GDP 441
 QY 503 ADPTAISGMKRRASDRV-----FDEPPCDPTMDLIEAKEKQAVFELCRRRYELQNP 555
 DB 442 SDTTRSPSVFTGKDNKVKKTRF-----DPSMEALGGQEFKPDLLIREDELREDIRNP 493

RESULT 6
 ID IF38_SCHPO STANDARD; PRT; 918 AA.
 AC O14164; P78791; Q9P6P4;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROBABLE EUKARYOTIC TRANSLATION INITIATION FACTOR 3 93 KDA SUBUNIT
 DE (EIF3 P93).
 GN TIF33 OR SPAC4A8.16C OR SPAC823.01C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Wood V.,
 RA Volckaert G., Aert R.;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 571-918 FROM N.A.
 RC STRAIN=PR745;
 RX MEDLINE=98162722; PubMed=9501991;
 RA Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;
 RT "Identification of open reading frames in Schizosaccharomyces pombe
 RT cDNAs.";
 RL DNA Res. 4:363-369(1997).
 CC -!- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF
 CC METHIONYL-TRNAI AND MRNA (BY SIMILARITY).
 CC -!- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 10 DIFFERENT SUBUNITS (BY
 CC SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE EIF3S8 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AL355013; CAB90146.1; -;
 DR EMBL; 298762; CAB11485.1; -;
 DR EMBL; D89140; BAA13802.1; -;
 DR InterPro: IPR000717; PCI.
 DR Pfam; PF01399; PCI; 1.
 DR SMART; SM00088; PINT; 2.
 KW Initiation factor; Protein biosynthesis.
 FT DOMAIN 1 188 ASP/GLU/SER-RICH.
 FT CONFLICT 683 683 M -> I (IN REF. 2).
 SQ SEQUENCE 918 AA; 104367 MW; 43B70599409688EF CRC64;

Query Match 4.7%; Score 145.5; DB 1; Length 918;
 Best Local Similarity 23.5%; Pred. No. 0.9;
 Matches 103; Conservative 57; Mismatches 172; Indels 107; Gaps 20;

[4] SEQUENCE OF 856-1940 FROM N.A.
 RP MEDLINE=8936648; PubMed=2771643;
 RA Karsch-Mirachi I., Travis M., Blau H., Leinwand L.A.;
 RT "Expression and DNA sequence analysis of a human embryonic skeletal
 muscle myosin heavy chain gene."
 RL Nucleic Acids Res. 17:6167-6179(1989).
 CC -!- FUNCTION: MUSCLE CONTRACTION.
 CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -!- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
 CC -!- DEVELOPMENTAL STAGE: ABUNDANTLY PRESENT IN FETAL SKELETAL MUSCLE
 CC AND NOT PRESENT OR BARELY DETECTABLE IN HEART AND ADULT SKELETAL
 CC MUSCLE.
 CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 CC MEROMYOSIN (LMN) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
 CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 CC SUBFRAGMENT (S2).
 CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X13988; CAA32167.1; -;
 DR EMBL; X13100; CAA31492.1; -;
 DR EMBL; X51593; CAA35942.1; -;
 DR EMBL; X15696; CAA33731.1; -;
 DR PIR; S04090; S04090.
 DR HSP; P13538; 2MYS.
 DR MIN; 160720; -;
 DR InterPro; IPR000048; IQ.
 DR InterPro; IPR002928; Myosin_tail.
 DR InterPro; IPR001609; myosin_head.
 DR Pfam; PF00612; IQ; 1.
 DR Pfam; PF00063; myosin_head; 1.
 DR Pfam; PF01576; Myosin_tail; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR ProDom; PD000355; myosin_head; 1.
 DR SMART; SM00015; IQ; 1.
 DR SMART; SM00242; MYSc; 1.
 DR PROSITE; PS50096; IQ; 1.
 DR Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;
 KW Multigene family.
 FT DOMAIN 1 781 MYOSIN HEAD-LIKE.
 FT DOMAIN 782 811 IQ.
 FT DOMAIN 840 1933 COILED COIL (POTENTIAL).
 FT NP_BIND 179 186 ATP (POTENTIAL).
 FT DOMAIN 656 678 ACTIN-BINDING.
 FT DOMAIN 758 772 ACTIN-BINDING.
 FT MOD_RES 130 130 METHYLATION (TR1-) (POTENTIAL).
 FT MOD_RES 696 696 ALKYLATION (SH-1).
 FT MOD_RES 706 706 ALKYLATION (SH-2).
 FT CONFLICT 1331 1331 A -> G (IN REF. 3).
 FT CONFLICT 1391 1392 KK -> QE (IN REF. 1 AND 2).
 FT CONFLICT 1608 1609 SR -> RA (IN REF. 3).
 FT CONFLICT 1663 1664 RG -> QT (IN REF. 2).
 SQ SEQUENCE 1940 AA; 224035 MW; 43CA58C6A4BA1253 CRC64;

Query Match

4.4%; Score 138; DB 1; Length 1940;

Best Local Similarity 19.6%; Pred. No. 5.7;
 Matches 109; Conservative 97; Mismatches 206; Indels 144; Gaps 25;
 QY 84 KKNETELANISGPHKKSTSTSTRKRARSKKK-ATDSVSDKIDESVASYDSS--THLRSS 141
 Db 941 KRKLEDECSL---KKDIDDLLETLAKYKEKHAHYENKVNKLTELSGLDTIAKLTR 996
 QY 142 SRSKKPVNVNSSSESEEQISKATKKVKQKEEVEEVDEKSLKNSSSDSEFEFVPE 201
 Db 997 KRALQEAHQALDQLAEDKVNLSNKTQKLEQQ--VEDL-ESSLEQE-----K 1043
 QY 202 QLETPISKRRSSSAKNLEKSTMNIDHAPREMPDCLDKPIPWGRGLGYACLAITLRS 261
 Db 1044 KLVVDLERNKRLKLGDKLQAESILDLENDKQQ-----LDERL-----K 1082
 QY 262 MKERVFCSRCTRIITQRDGL---SVKQLGTQNVLDLKLVEWNNHFIHFMVYSSDLF 318
 Db 1083 KDFEYVCQLOSKVEDRDTGLQFOKKIKELQAR--IELEEEIEAER----- 1127
 QY 319 PFASHAK-----YCYTLEFAQSHLEEVGKLANKNHRLTMHPGOYTQIA--SPREVV 369
 Db 1128 --ATRAKTEKORSYARELEELSERLEEAAGVTS-----TQIELNKKREAEF 1172
 QY 370 DSAIRDL---AYHDEILSKMKLNQNLKDAVLIHLGTFEGKKTLDLRFKNYORLSD 425
 Db 1173 LKLRRDLEATLOHEAMVATLR-----KKHDSVAELG-----EODNLQORVKKL-E 1219
 QY 426 SVKARLVLENDVMSVQ-----DLLPLCQELATPLVLDWHHHNIVPGTLREGSLDM 478
 Db 1220 KKKSEFKLEITDOLSSSMESVSKANLEKICRTLEDQL-----SEARGKNEEQRSLS 1272
 QY 479 PLIPTIRETWTRKGIQKOHYSESADPTAISGMKRAHSRDVDFPDCPTMDLMIEAKE 538
 Db 1273 ELTOKSRLOTAGELSRLEEKESIVSOLS-----RSKQAF-----TQOTEELKRLQEE 1322
 QY 539 KEQAVFELCRRYELQNPCCLEIMGPEYDQTRDGYPPGCAEKRLTARKRS----- 589
 Db 1323 ENKAKNALAHALQSSRHDC--DLLREQVEEQEG-----KAEQLRALSKANSEVAQWRTKY 1376
 QY 590 -----RKEEVEDEK 599
 Db 1377 EYDATORTEEELEAKK 1392
 RESULT 12
 MSP1_PLAFK
 ID MSP1_PLAFK STANDARD; PRT: 1630 AA.
 AC P04932;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
 DE (PMMSA) (P190).
 GN MSP-1.
 OS Plasmodium falciparum (isolate K1 / Thailand).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5839;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=86136024; PubMed=3004972;
 RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
 RA Stunnenberg H., Bujard H.;
 RT "Polymorphism of the precursor for the major surface antigens of
 RL Plasmodium falciparum merozoites: studies at the genetic level.";
 RN EMBO J. 4:3823-3829(1985).
 RN [2]
 RP REVISIONS, SEQUENCE FROM N.A.
 RA Pan W., Tolle R., Bujard H.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 CC (POTENTIAL).
 CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
 CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF

FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1174 1174 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1445 1445 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1526 1526 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1639 AA; 187618 MW; 2C255B6616C87F6E CRC64;

Query Match 4.4%; Score 136.5; DB 1; Length 1639;
 Best Local Similarity 19.5%; Pred. No. 5.6;
 Matches 130; Conservative 107; Mismatches 212; Indels 217; Gaps 35;

QY 17 FTILKQAFKGNHPCVSPVCTIYS---RFHCLPTLKLSPMSKTTLSMLPQVNTIGAN 73
 Db 185 FOLLRAKL---NDVCANDYQCPFLKTRANEL-DVLKLV----- 221

QY 74 SFSAETPVDLKKENETELANISGPHKKSTSTRKRARSSKKKADTSVDKIDESVASYD 133
 Db 222 -FGYRKPLNDKNDGVKMDYI---KKN-----KKTIENTINELIESKTTID 264

QY 134 SSTHLRRSRSKPPV-----YNSSESESEEOISATKKVKOKKEEEYVEEYDEK-- 184
 Db 265 KKNKATKEEKKKLYQAOYLSYINKOLE-EAHNLISVLEKRIIDTLKKNENIKELLDKIN 323

QY 185 SLKN-----ESSDEREPVPEQLET-----PISKRERSRSA 217
 Db 324 EIKNPPANGSNTPNTLLDKNNKIEHEKEKEIAKTIKFNIDSLFTDPLEYLYLREKN 383

QY 218 KNLE-----KESPMNLDHAPREMFCDLKPWPWRGLGYACNLITLISM----- 262
 Db 384 KNIDISAKVETKEST-----EPNEYPNGVTYP-----LSYNDINNALNELNFGDLIN 431

QY 263 -----KERVCSRTCRTITTTQDGLSVKQL---GTQNVLDLKLVE--- 301
 Db 432 PDYTKPSKNIYTDNERKKFTNEIKIEKKKIESDKRSYEDRSKSLNDITKEYL 491

QY 302 -----WNHFGIHEMRVSSDLFPFAS--HAKYGYTFLEPAQSHLEVGKLANKNHRL 351
 Db 492 LNEIYDSKFNNI-----DLTNEFKMMGKRYSYKVEKLTHNTFASVENSKHLEK 542

QY 352 TMHPQOYTQIASPREVVYDSAIROLAYHDEILSRMK-----LNEOLNKDAVLIHLGGTF 406
 Db 543 LTKALKYMEDYSLRNIVVE---KELKYKNLISKIENBIETLVENIKKDEQL-----F 593

QY 407 EGKKTFLDRFRKNYQ--RLSDSVKAR-----LVLENDVSVSVQDQLPLCQELNIPVLVDW 460
 Db 594 E-KKTKDENPDEKILEVSDIVKQVQVLLMMNKIDELKKTQILK-----NVEL----- 643

QY 461 HHHNT-VPGTLREGS-----LD-----IMPLIPTIRETWTRKGIQKHYSSES 502
 Db 644 -KHNIHVPNSYKQENKQEPYVYLLVLKKEIDKLKVPKVESLNEEKKNKTKTEGQ--SDN 700

QY 503 ADPTA---ISGMKRRASDRYDFDPPCDPTMDLMEAKEKEQAVFELCRRYELONPPCPL 559
 Db 701 SEPTGEITQATTKPGQAGSALBGD---SVOAQAOEQKQA-----QPPVPV 746

QY 560 EIMGPE 565
 Db 747 PV--PE 750

RESULT 14

KELI_YEAST
 ID KELI_YEAST STANDARD; PRT; 1164 AA.
 AC P38853;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE KELCH REPEATS PROTEIN 1.
 GN KELI OR YHR158C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;

Query Match 4.3%; Score 135; DB 1; Length 1164;
 Best Local Similarity 21.6%; Pred. No. 4.4;
 Matches 84; Conservative 64; Mismatches 147; Indels 94; Gaps 16;

QY 62 LSMLPOVNLGANSFSAETPVDLKKENETELANISGPHKKSTSTRKRARSSKKKATDSV 121
 Db 616 LEVLPSEI-----KTP---QNCNIEIKHLPDADEKIDSTTTTDFQEIINGDKLGTSSM 665

QY 122 SDKIDE--SVASYDSSTHLR---SSRSK---KPVNYSNSESSESEEOISATKKVKOKE 173
 Db 666 S-KVEEDGNVADEDEIGVAQWASSPSKQDFKIKHYNESSELSONNTEIDKLS----- 717

QY 174 EEEYVEEVEDEKSLKNESSE-DEFEVPVPEQLETPIKSRRRSRSSAKNLEKSTMMNLDHHA 232
 Db 718 -----EPVDITTKSDTAGHDSANHVIDASDEKNVSPMGVDPTDKNEEASVPINRD--A 770

QY 233 PREMFCDLKDPIPWGRGLCYACNLITLRSKMERVFCSTRTITTTQDGLSVKOLGTGN 292
 Db 771 TTEVVD-----RALPEKL-----RSELQSLKELTHEK 797

QY 293 VLIDLKLVENHNFGIHFMVRVSSDLFPFASHAKYGYTFLEFAQ---SHLEEVGKLANKNYH 349
 Db 293 VLIDLKLVENHNFGIHFMVRVSSDLFPFASHAKYGYTFLEFAQ---SHLEEVGKLANKNYH 349

[1]

RP SEQUENCE FROM N.A.
 RP STRAIN-S288C / AB972;
 RC MEDLINE=94378003; PubMed=8091229;
 RX Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
 RA Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,
 RA Kuback T., Hillier L., Jier M., Johnston L., Langston Y.,
 RA Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,
 RA Nham M., Rifkin L., Riles L., St. Peter H., Trevaskis E., Vaughan K.,
 RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
 RA Vaudin M.;
 RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
 RT VIII.";
 RL Science 265:2077-2082(1994).
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE=99003296; PubMed=9786949;
 RA Phillips J., Herskowitz I.;
 RT "Identification of Kel1p, a kelch domain-containing protein involved
 RT in cell fusion and morphology in Saccharomyces cerevisiae.";
 RL J. Cell Biol. 143:375-389(1998).
 CC -!- FUNCTION: HAS A ROLE IN CELL MORPHOGENESIS AND CELL FUSION AND MAY
 CC ANTAGONIZE THE PKC1 PATHWAY.
 CC -!- SUBUNIT: INTERACTS WITH KEL2.
 CC -!- SIMILARITY: CONTAINS 5 KELCH REPEATS.
 CC -!- SIMILARITY: TO YEAST KEL2.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: U10397; AAB68991.1; --
 DR PIR: S46769; S46769.
 DR SGD: S0001201; KEL1.
 DR InterPro: IPR001798; Kelch.
 DR Pfam: PF01344; Kelch; 3.
 KW Repeat; Coiled coil.
 FT REPEAT 139 186
 FT REPEAT 253 307
 FT REPEAT 308 357
 FT REPEAT 359 409
 FT REPEAT 411 460
 FT REPEAT 477 531
 FT DOMAIN 974 1163
 FT COILED COIL (POTENTIAL).
 FT COILED COIL (POTENTIAL).
 SQ SEQUENCE 1164 AA; 131093 MW; 43D0FC570F1D5E4D CRC64;

Db 798 ALEA-----GAHIKELETELWOLKSOQNSGTTKEIDELDSVRLOSKEILLEADNH 847
Qy 350 RUTMHPGQVQTASPREVVVDSAIRDLAYHDEILSRMKLNEOLNKDAVLIHLGGTFEGK 409
Db 848 SL-----EDKVNELEELVNSKFLDIENLEVI-----QFQNEKIKSLELEPNYREK 893
Qy 410 KETLDRFRKNYQRLSDSVKARLVLENDV 438
Db 894 LEELOIHEHNLRENERLKNESKQHNEI 922

RESULT 15

POLR_ASCVP STANDARD; PRT: 2105 AA.
AC P36309;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: RNA REPLICASE (EC 2.7.7.48); HELICASE;
DE COAT PROTEIN].
OS Apple stem grooving virus (strain P-209) (ASGV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Capilliovirus.
OX NCBI_TaxID=36402;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93033164; PubMed=1413530;
RA Yoshikawa N., Sasaki E., Kato M., Takahashi T.;
RT "The nucleotide sequence of apple stem grooving capilliovirus genome.";
RL Virology 191:98-105(1992).
CC 1- PTM: THE N-TERMINAL OF THE COAT PROTEIN IS BLOCKED.
CC 1- MISCELLANEOUS: THE COAT PROTEIN IS LOCATED IN THE CARBOXY-TERMINAL
CC REGION OF THIS POLYPROTEIN.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

CC EMBL; D14995; BAA03639.1; -
CC PIR; A44059; A44059.
CC MEROPS; C35.001; -
KW RNA-directed RNA polymerase; Transferase; Polyprotein; ATP-binding;
KW Coat protein; Helicase.
FT NP_BIND 781 788 ATP (POTENTIAL).
FT DOMAIN 1364 1453 POLYMERASE ACTIVE SITE.
SQ SEQUENCE 2105 AA; 241240 MW; 331291A5FA137131 CRC64;

Query Match 4.3%; Score 134; DB 1; Length 2105;
Best Local Similarity 21.2%; Pred. No.10;
Matches 113; Conservative 81; Mismatches 193; Indels 146; Gaps 28;

Qy 119 DSVSDKIDESVAVSDSTHLRRS-----SRSKKP-----VNVNNSSESESER 160
Db 491 NSGLKIDYSLVRRFRRLRDFLDIPKSGSPRKQFLFLESTGNKSNPNAEKNSGES 550
Qy 161 QISKATKKVKQKEEYVEVEDEKSLKNSSDDEFVVPPEQL-----ETPISKRRRS 213
Db 551 -----IKIEGSAENDQPHVEVSHTSMTEDQG-FEGSIPVDLINCPEPEIKLPKRRK 603
Qy 214 R-----SSAKNLEKSTWN--LDDHAPREMFCDLKPWPGRGLGYACNLNLTIL----- 259
Db 604 NDCVFKAISAHLGIDSQDLLNLFVNEIDISDELLDCTEED---KG-LSHEMIEEVLITKGL 659
Qy 260 -----RSMKE-----RVFC-----SRTCRITTIORDG--LESVKQLGT 290
Db 660 SMVYTSDFKEMAVLNRYGVNGMYCTIKGNHCELSKCEFIKLLKEGGAQMSNENLNA 719
Qy 291 QNVLDLIKLVENHN-----FGIHF---MRVSSDLFPFASHAKY 326
:: || | | |

Search completed: January 15, 2002, 14:07:05
Job time: 934 sec

Db 720 DSLFDLGRFV---HNRDRAVKLAKSMARGTTGLLNEFDLEFCKNMVTLSELFPEPFSSVV 776
Qy 327 CYTLEPAQSHLEEVGKLANKNYHRLTMHPGOYTOIASPREVVVDSAIRDLAYHDEILSRM 386
Db 777 GLRLGFAGS-----GK-THKVLQWINTPYSVKRMFISPRMLA-----DEVEPOL 820
Qy 387 KLNE-OLNKDAVLIHLGGTF-EGKKTLDORFRKNYQRLSDSVKARLVLENDVSVSWQD 444
Db 821 KGTACQVHTWETALKKIDGTGFMFVDEIGLYPPGYLTLLQMCFAFRKIVKQSENFLKKG 880
Qy 445 LLPLCQE-LNI-----PLVLDWHHHNVPGTLREGSLDLPLPIPTIRETWTRKGIQKH 498
Db 881 LLELSKTCLNIRCFCDPLQL--RYSAEDTNLLDKTHDIDLMIKTIKHKYLFQYRGQW 938
Qy 499 YESAD-PTAI--SGMKRRASHD----RVDFPPCDPTMDLMIEAKKEQAVF 544
Db 939 FOELVNMPTRVDESFSKFKFFADISSVKTEY-----GLILVAKREDKGVF 984

```

Db 63 VKLTQSMALIRYIADKHNMGLGCPKRAEISMLEGAVLDIRYGVSRITAYSKDFETLKVD 122
QY 125 LSKLPEMLKMFEDRLCHKTYLNGDHTVHPDMLYDALDVVLYMDPCLDAPFKLVCFKKR 184
Db 123 LSKLPEMLKMFEDRLCHKTYLNGDHTVHPDMLYDALDVVLYMDPCLDAPFKLVCFKKR 182
QY 185 IEAIPQIDKYLKSSKYIAMPLOGWQATFGGDDHPPKSDHLPVPRG--SMRLRLKRNQIISK 242
Db 183 IEAIPQIDKYLKSSKYIAMPLOGWQATFGGDDHPPKSD--LVPRGSPSLOQAENNTITNSK 241
QY 243 RIVFTILKQAFKGNHPCVPSVCTIT-----YSRFHCLPD-----TLKSLLPMSST 289
Db 242 K-EMTKLREKVKKAEKEDLAINRATKLEERNQAYKAAHKAEEBKAKTFQRLITFESEN 300
QY 290 -TSLMLPQVNI-----ANSFSAETPVDLKKENETELANISGPHKKSTSTSTRKRARSK 343
Db 301 INLKRPNDVAVSNRDKKNSSETAKTD-EVEKQRAAEAAKAVETEKORAAEAT-KVAEAEK 358
QY 344 KKATDSV-----SDKIDESVASYDSSTHLRSSRSKPKPVNYSNSESSESEEQISKATKKVK 399
Db 359 RKAEEAAKAVETEKORAAEATKVAEAEKQAAEAAKAV-----ETEKORAAEATKVA 410
QY 400 QKEEEYVEEVDEKSLKNSSSDPEFVVPVPEQLETPISKRRRSRSAAKNLEKE 452
Db 411 EAEKQRAAEAMKVAEAEKQAA-----EATKVAEAEKQAAEATKVAEAE 455

```

RESULT 2

```

US-09-115-746-2
; Sequence 2, Application US/09115746
; Patent No. 6228601
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYANOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115.746
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/216,894
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 564 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-115-746-2

```

Query Match 29.0%; Score 1259.5; DB 4; Length 564;

```

Best Local Similarity 57.1%; Pred. No. 1.9e-94;
Matches 270; Conservative 55; Mismatches 103; Indels 45; Gaps 12;
QY 5 PILGYWKIKGLVQTRILLLEYLBEEKYEEHLYERDEGDKWRKKFELGLEFPNLPYYIDGD 64
Db 3 PILGYWKIKGLVQTRILLLEYLBEEKYEEHLYERDEGDKWRKKFELGLEFPNLPYYIDGD 62
QY 65 VKLTQSMALIRYIADKHNMGLGCPKRAEISMLEGAVLDIRYGVSRITAYSKDFETLKVD 124
Db 63 VKLTQSMALIRYIADKHNMGLGCPKRAEISMLEGAVLDIRYGVSRITAYSKDFETLKVD 122
QY 125 LSKLPEMLKMFEDRLCHKTYLNGDHTVHPDMLYDALDVVLYMDPCLDAPFKLVCFKKR 184
Db 123 LSKLPEMLKMFEDRLCHKTYLNGDHTVHPDMLYDALDVVLYMDPCLDAPFKLVCFKKR 182
QY 185 IEAIPQIDKYLKSSKYIAMPLOGWQATFGGDDHPPKSDHLPVPRG--SMRLRLKRNQIISK 242
Db 183 IEAIPQIDKYLKSSKYIAMPLOGWQATFGGDDHPPKSD--LVPRGSPSLOQAENNTITNSK 241
QY 243 RIVFTILKQAFKGNHPCVPSVCTIT-----YSRFHCLPD-----TLKSLLPMSST 289
Db 242 K-EMTKLREKVKKAEKEDLAINRATKLEERNQAYKAAHKAEEBKAKTFQRLITFESEN 300
QY 290 -TSLMLPQVNI-----ANSFSAETPVDLKKENETELANISGPHKKSTSTSTRKRARSK 343
Db 301 INLKRPNDVAVSNRDKKNSSETAKTD-EVEKQRAAEAAKAVETEKORAAEAT-KVAEAEK 358
QY 344 KKATDSV-----SDKIDESVASYDSSTHLRSSRSKPKPVNYSNSESSESEEQISKATKKVK 399
Db 359 RKAEEAAKAVETEKORAAEATKVAEAEKQAAEAAKAV-----ETEKORAAEATKVA 410
QY 400 QKEEEYVEEVDEKSLKNSSSDPEFVVPVPEQLETPISKRRRSRSAAKNLEKE 452
Db 411 EAEKQRAAEAMKVAEAEKQAA-----EATKVAEAEKQAAEATKVAEAE 455

```

RESULT 3

```

US-08-216-894-8
; Sequence 8, Application US/08216894
; Patent No. 5876734
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYANOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/216,894
; FILING DATE: 24-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 643 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```



```

; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 354 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-216-894-4

```

```

Query Match      28.2%; Score 1228.5; DB 2; Length 354;
Best Local Similarity 69.3%; Pred. No. 3.1e-92;
Matches 250; Conservative 24; Mismatches 60; Indels 27; Gaps 7;

```

```

QY 5 PILGYWKIKGLVQPTRLLEYLEEKYEHLRYERDEGDKWRNKKFELGLEFPNLYYIDGD 64
   |||||
Db 3 PILGYWKIKGLVQPTRLLEYLEEKYEHLRYERDEGDKWRNKKFELGLEFPNLYYIDGD 62
   |||||
QY 65 VKLTOSMAIIRIADKHNLGCGPKERAIEISMLEGAVLDIRYGVSRAYSKDFTETLKVD 124
   |||||
Db 63 VKLTOSMAIIRIADKHNLGCGPKERAIEISMLEGAVLDIRYGVSRAYSKDFTETLKVD 122
   |||||
QY 125 LSKLEPMLKMFEDRLCHTKYLNQDHTVHPDMLYDALDVLYMDPCLDAPPKLVCFKKR 184
   |||||
Db 123 LSKLEPMLKMFEDRLCHTKYLNQDHTVHPDMLYDALDVLYMDPCLDAPPKLVCFKKR 182
   |||||
QY 185 IEAIPQIDKYLKSSKYIAWPLQGWQATFGGDHPPKSDHLPVPG--SMRLRLKRNIOISK 242
   |||||
Db 183 IEAIPQIDKYLKSSKYIAWPLQGWQATFGGDHPPKSD--LVPRSPSOLQQAENNIINSK 241
   |||||
QY 243 RIVFTILKOKAFKGNHPCVPSVCTIT-----YSRFHCLPD-----TLKSLPMSKKT 289
   |||||
Db 242 K-EMTKLEKVKKAEEKLDINRATKLEERNQAYKAAHKAEEKAKTFORLITFESE- 299
   |||||
QY 290 TSLMLPQVNIANSFSAETPVDLKENETELANISGPHKSTSTSTRKARSKKKATDS 349
   |||||
Db 300 -----NINLKKRPNDVSNRDKKNSETAKTD-----EVEKQRAAAKAVETEKORAGEF 350
   |||||
QY 350 V 350
Db 351 I 351

```

```

RESULT 6
US-09-115-746-4
; Sequence 4, Application US/09115746
; Patent No. 6228601
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115.746

```

```

; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/216,894
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 354 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-115-746-4

```

```

Query Match      28.2%; Score 1228.5; DB 4; Length 354;
Best Local Similarity 69.3%; Pred. No. 3.1e-92;
Matches 250; Conservative 24; Mismatches 60; Indels 27; Gaps 7;

```

```

QY 5 PILGYWKIKGLVQPTRLLEYLEEKYEHLRYERDEGDKWRNKKFELGLEFPNLYYIDGD 64
   |||||
Db 3 PILGYWKIKGLVQPTRLLEYLEEKYEHLRYERDEGDKWRNKKFELGLEFPNLYYIDGD 62
   |||||
QY 65 VKLTOSMAIIRIADKHNLGCGPKERAIEISMLEGAVLDIRYGVSRAYSKDFTETLKVD 124
   |||||
Db 63 VKLTOSMAIIRIADKHNLGCGPKERAIEISMLEGAVLDIRYGVSRAYSKDFTETLKVD 122
   |||||
QY 125 LSKLEPMLKMFEDRLCHTKYLNQDHTVHPDMLYDALDVLYMDPCLDAPPKLVCFKKR 184
   |||||
Db 123 LSKLEPMLKMFEDRLCHTKYLNQDHTVHPDMLYDALDVLYMDPCLDAPPKLVCFKKR 182
   |||||
QY 185 IEAIPQIDKYLKSSKYIAWPLQGWQATFGGDHPPKSDHLPVPG--SMRLRLKRNIOISK 242
   |||||
Db 183 IEAIPQIDKYLKSSKYIAWPLQGWQATFGGDHPPKSD--LVPRSPSOLQQAENNIINSK 241
   |||||
QY 243 RIVFTILKOKAFKGNHPCVPSVCTIT-----YSRFHCLPD-----TLKSLPMSKKT 289
   |||||
Db 242 K-EMTKLEKVKKAEEKLDINRATKLEERNQAYKAAHKAEEKAKTFORLITFESE- 299
   |||||
QY 290 TSLMLPQVNIANSFSAETPVDLKENETELANISGPHKSTSTSTRKARSKKKATDS 349
   |||||
Db 300 -----NINLKKRPNDVSNRDKKNSETAKTD-----EVEKQRAAAKAVETEKORAGEF 350
   |||||
QY 350 V 350
Db 351 I 351

```

```

RESULT 7
US-08-216-894-10
; Sequence 10, Application US/08216894
; Patent No. 5876734
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```



```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/216,894
; FILING DATE: 24-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-216-894-10

```

```

Query Match      28.1%; Score 1221.5; DB 2; Length 472;
Best Local Similarity 56.7%; Pred. No. 1.8e-91;
Matches 255; Conservative 54; Mismatches 110; Indels 31; Gaps 9;

Qy  5 PILGYWKIKGLVQTRLLLEYLEEYEHLYERDEGDKWRNKKFELGLFPNLPYYIDGD 64
Db  3 PILGYWKIKGLVQTRLLLEYLEEYEHLYERDEGDKWRNKKFELGLFPNLPYYIDGD 62

Qy  65 VKLTQSMALIRYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRISYASKDFETLKVD 124
Db  63 VKLTQSMALIRYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRISYASKDFETLKVD 122

Qy  125 LSKLPEMLKMPEDRLCHTKYTLNGDHVTHPDVLMYDALDVVLYMDPMDCLDAFPKLVCFK 184
Db  123 LSKLPEMLKMPEDRLCHTKYTLNGDHVTHPDVLMYDALDVVLYMDPMDCLDAFPKLVCFK 182

Qy  185 IEAIPQIDKYLKSSKYIAMPLOGHQATFGGDDHPPKSDHLVPRGSMRLRLKRNIIQISKRI 244
Db  183 IEAIPQIDKYLKSSKYIAMPLOGHQATFGGDDHPPKSDHLVPRGSMRLRLKRNIIQISKRI 237

Qy  245 VFTILKQAFKGNHPCVPVCTIYTSRHFCLPDTLKSLLPMSSKTTLSMLPOVNIANSF 304
Db  238 A-EAEKQAAEAT-----KVAEAEKQAAEATK--VAEAEKQAAEATKVAEAEKQ 286

Qy  305 SAETP--VDLKKENETELANTISGPHKSTSTSTRKRARRSKKATDSVSKIDESVASYD 362
Db  287 AAEATKVAEAEKQAAEATKVAGDEKQAAEAT--KVAEAEKQAAEAT--KVAE--AEKQ 341

Qy  363 SSTHLRRSSRSKPKVNYNSSESESEBQISKATKKVKQKEEEYVEEVEDEKSLKNSSSD 422
Db  342 KAAEATKVAEAEKQAAEATKVAEAEKQAAEATKVAEAEKQAAEATKVAEAEKQAA- 400

Qy  423 EFEPVVPQLETPISKRRSSRSKAKNLEKE 452
Db  401 -----EATKVAEAEKQAAEATKVAEAE 423

```

```

RESULT 8
US-09-115-746-10
; Sequence 10, Application US/09115746
; Patent No. 6228601
; GENERAL INFORMATION:
; APPLICANT: Kirchoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.

```

```

; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115,746
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/216,894
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-115-746-10

```

```

Query Match      28.1%; Score 1221.5; DB 4; Length 472;
Best Local Similarity 56.7%; Pred. No. 1.8e-91;
Matches 255; Conservative 54; Mismatches 110; Indels 31; Gaps 9;

Qy  5 PILGYWKIKGLVQTRLLLEYLEEYEHLYERDEGDKWRNKKFELGLFPNLPYYIDGD 64
Db  3 PILGYWKIKGLVQTRLLLEYLEEYEHLYERDEGDKWRNKKFELGLFPNLPYYIDGD 62

Qy  65 VKLTQSMALIRYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRISYASKDFETLKVD 124
Db  63 VKLTQSMALIRYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRISYASKDFETLKVD 122

Qy  125 LSKLPEMLKMPEDRLCHTKYTLNGDHVTHPDVLMYDALDVVLYMDPMDCLDAFPKLVCFK 184
Db  123 LSKLPEMLKMPEDRLCHTKYTLNGDHVTHPDVLMYDALDVVLYMDPMDCLDAFPKLVCFK 182

Qy  185 IEAIPQIDKYLKSSKYIAMPLOGHQATFGGDDHPPKSDHLVPRGSMRLRLKRNIIQISKRI 244
Db  183 IEAIPQIDKYLKSSKYIAMPLOGHQATFGGDDHPPKSDHLVPRGSMRLRLKRNIIQISKRI 237

Qy  245 VFTILKQAFKGNHPCVPVCTIYTSRHFCLPDTLKSLLPMSSKTTLSMLPOVNIANSF 304
Db  238 A-EAEKQAAEAT-----KVAEAEKQAAEATK--VAEAEKQAAEATKVAEAEKQ 286

Qy  305 SAETP--VDLKKENETELANTISGPHKSTSTSTRKRARRSKKATDSVSKIDESVASYD 362
Db  287 AAEATKVAEAEKQAAEATKVAGDEKQAAEAT--KVAEAEKQAAEAT--KVAE--AEKQ 341

Qy  363 SSTHLRRSSRSKPKVNYNSSESESEBQISKATKKVKQKEEEYVEEVEDEKSLKNSSSD 422
Db  342 KAAEATKVAEAEKQAAEATKVAEAEKQAAEATKVAEAEKQAAEATKVAEAEKQAA- 400

Qy  423 EFEPVVPQLETPISKRRSSRSKAKNLEKE 452
Db  401 -----EATKVAEAEKQAAEATKVAEAE 423

```

```

RESULT 9
US-08-395-507-2
; Sequence 2, Application US/08395507
; Patent No. 5578456
; GENERAL INFORMATION:
; APPLICANT: Fujimura, Katsuya

```



```
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..515
; OTHER INFORMATION: /note= "fusion protein composed of
; OTHER INFORMATION: glutathione-S-transferase sequence,
; OTHER INFORMATION: thrombin cleavage sequence, recognition
; OTHER INFORMATION: sequence for heart muscle protein
; OTHER INFORMATION: kinase, residues introduced by cloning
; OTHER INFORMATION: and hprt protein fragment"
; US-08-974-549A-604

Query Match 27.7%; Score 1204.5; DB 4; Length 515;
Best Local Similarity 69.7%; Pred. No. 5e-90;
Matches 241; Conservative 17; Mismatches 33; Indels 55; Gaps 5;

QY 5 PILGYWKIKGLVQPTRLLEYLEEYEEHLYERDEGDKWRNKKFELGFPNLPYYIDG 64
Db 3 PILGYWKIKGLVQPTRLLEYLEEYEEHLYERDEGDKWRNKKFELGFPNLPYYIDG 62
QY 65 VKLTQSMALIRYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRVAYSKDFETLKVD 124
Db 63 VKLTQSMALIRYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRVAYSKDFETLKVD 122
QY 125 LSKLPEMLKMFEDRLCHKTYLNGDHTVTHPDFMLYDALDVVLYMDPCLDAFFPKLVCFKR 184
Db 123 LSKLPEMLKMFEDRLCHKTYLNGDHTVTHPDFMLYDALDVVLYMDPCLDAFFPKLVCFKR 182
QY 185 IEAIPQIDKYKSKYIAWPLQGWQATFGGDDHPPKSDHLPVGRS-----M 230
Db 183 IEAIPQIDKYKSKYIAWPLQGWQATFGGDDHPPKSD-LVPRGSRASVGSVTYKMSYV 241
QY 231 LRLKRNQIS-----KRIVTILKOKAFKGNHPCV 261
Db 242 VELURFFVYVTTFOKNLFFRPSWWSKLSQIGIRQHLKRVQLRELSEAEVRQHREAR 301
QY 262 PSYCTIYSRFLCPDLTKSLPLMSKTTLSMLPQVNICANSFSAE 307
Db 302 PALLT-SRLRFIPKDGRLPVM-----DYVVGARTFERE 336

RESULT 13
US-08-974-549A-605
; Sequence 605, Application US/08974549A
; Patent No. 6166178
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin B.
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: Human Telomerase Catalytic Subunit
; NUMBER OF SEQUENCES: 727
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,549A
```

```
; FILING DATE: 19-NOV-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/854,050
; FILING DATE: 09-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/911,312
; FILING DATE: 14-AUG-1997
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/912,951
; FILING DATE: 14-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/915,503
; FILING DATE: 14-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/17618
; FILING DATE: 01-OCT-1997
; APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/17885
; FILING DATE: 01-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph Ted
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-002610US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 605:
; LENGTH: 514 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..515
; OTHER INFORMATION: /note= "fusion protein composed of
; OTHER INFORMATION: glutathione-S-transferase sequence,
; OTHER INFORMATION: thrombin cleavage sequence, recognition
; OTHER INFORMATION: sequence for heart muscle protein
; OTHER INFORMATION: kinase, residues introduced by cloning
; OTHER INFORMATION: and hprt protein fragment"
; US-08-974-549A-605
```

```
Query Match 27.7%; Score 1203; DB 4; Length 514;
Best Local Similarity 71.6%; Pred. No. 6.6e-90;
Matches 244; Conservative 13; Mismatches 36; Indels 48; Gaps 7;

QY 5 PILGYWKIKGLVQPTRLLEYLEEYEEHLYERDEGDKWRNKKFELGFPNLPYYIDG 64
Db 3 PILGYWKIKGLVQPTRLLEYLEEYEEHLYERDEGDKWRNKKFELGFPNLPYYIDG 62
QY 65 VKLTQSMALIRYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRVAYSKDFETLKVD 124
Db 63 VKLTQSMALIRYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRVAYSKDFETLKVD 122
QY 125 LSKLPEMLKMFEDRLCHKTYLNGDHTVTHPDFMLYDALDVVLYMDPCLDAFFPKLVCFKR 184
Db 123 LSKLPEMLKMFEDRLCHKTYLNGDHTVTHPDFMLYDALDVVLYMDPCLDAFFPKLVCFKR 182
```

QY 185 IEAIPQIDKYLKSSKYIAWPLQGWQATFGGDPKSDHLYVPRGSMRLRLKRNIOISKRI 244
DB 183 IEAIPQIDKYLKSSKYIAWPLQGWQATFGGDPKSDHLYVPRGSMRLRLKRNIOISKRI 230
QY 245 VFTILKOKAFKG-----NHPKV-----PSV-----CTITYSRF-HCL----- 275
DB 231 VGSVTKATSEALSGSTRSHPSVGRHAGPPSSRPPRPWDTPCPPVYATKHFLYSS 290
QY 276 -----POTLSLPLMSKTTLSMLPQVNICANFSFAETP 309
DB 291 GQKEQLRPSFLLSLRPSLTGARRLVETIFLGRPMWPGTP 331

RESULT 14

US-08-395-507-1
; Sequence 1, Application US/08395507
; Patent No. 5578456
; GENERAL INFORMATION:
; APPLICANT: Fujimura, Katsuya
; APPLICANT: Ueno, Eichi
; APPLICANT: Fujii, No. 5578456uyuki
; APPLICANT: Okada, Masahisa
; TITLE OF INVENTION: Anti-treponema Pallidum Antibody
; TITLE OF INVENTION: Immunoassay
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pollock, Vande Sande & Priddy
; STREET: 1990 M Street, NW, Suite 800
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/395,507
; FILING DATE: 27-FEB-1995
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP Hei-6-54672
; FILING DATE: 28-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Amernick, Burton A.
; REGISTRATION NUMBER: 24,852
; REFERENCE/DOCKET NUMBER: 1581/007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 331-7111
; TELEFAX: (202) 293-6229
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 352 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: rabbit
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..352
; OTHER INFORMATION: /note= "G15 Antigen"
US-08-395-507-1

Query Match 27.6%; Score 1201.5; DB 1: Length 352;
Best Local Similarity 99.1%; Pred. No. 5e-90;
Matches 225; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 5 PILGYWKIGLVQPTRLLEYLEEKYEHELYERDEGDKWRNKKFELGLEFPNLPYYIDGD 64
DB 3 PILGYWKIGLVQPTRLLEYLEEKYEHELYERDEGDKWRNKKFELGLEFPNLPYYIDGD 62
QY 65 VKLTQSMATIRYIADKINMLGGCPKRAEISMLEGAVLDIRYGVSRITAYSKDFETLKVDF 124
DB 63 VKLTQSMATIRYIADKINMLGGCPKRAEISMLEGAVLDIRYGVSRITAYSKDFETLKVDF 122
QY 125 LSKLPEMLKMFEDRLCHTKYTLNGDHTVDPFMYLDALDVLVYMDPMLCDAPFKLVCFKKR 184
DB 123 LSKLPEMLKMFEDRLCHTKYTLNGDHTVDPFMYLDALDVLVYMDPMLCDAPFKLVCFKKR 182
QY 185 IEAIPQIDKYLKSSKYIAWPLQGWQATFGGDPKSDHLYVPRGSMRL 231
DB 183 IEAIPQIDKYLKSSKYIAWPLQGWQATFGGDPKSDHLYVPRGSMRL 228
RESULT 15
US-08-974-549A-603
; Sequence 603, Application US/08974549A
; Patent No. 6166178
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin B.
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: Human Telomerase Catalytic Subunit
; NUMBER OF SEQUENCES: 727
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,549A
; FILING DATE: 19-NOV-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/854,050
; FILING DATE: 09-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/911,312
; FILING DATE: 14-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/912,951
; FILING DATE: 14-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/915,503
; FILING DATE: 14-AUG-1997
; PRIOR APPLICATION DATA:

/ APPLICATION NUMBER: WO PCT/US97/17618
/ FILING DATE: 01-OCT-1997
/ PRIOR APPLICATION DATA: WO PCT/US97/17885
/ APPLICATION NUMBER: WO PCT/US97/17885
/ FILING DATE: 01-OCT-1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Apple, Randolph Ted
/ REGISTRATION NUMBER: 36,429
/ REFERENCE/DOCKET NUMBER: 015389-002610US
/ TELEPHONE: (415) 576-0200
/ TELEFAX: (415) 576-0300
/ INFORMATION FOR SEQ ID NO: 603:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 530 amino acids
/ TYPE: amino acid
/ STRANDEDNESS:
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ FEATURE:
/ NAME/KEY: Protein
/ LOCATION: 1..530
/ OTHER INFORMATION: /note= "fusion protein composed of
/ OTHER INFORMATION: glutathione-S-transferase sequence,
/ OTHER INFORMATION: thrombin cleavage sequence, recognition
/ OTHER INFORMATION: sequence for heart muscle protein
/ OTHER INFORMATION: kinase, residues introduced by cloning,
/ OTHER INFORMATION: eight consecutive His residues and hTBT
/ OTHER INFORMATION: protein fragment"
US-08-974-549A-603

Search completed: January 15, 2002, 13:50:57
Job time: 326 sec

Query Match 27.6%; Score 1200.5; DB 4; Length 530;
Best Local Similarity 50.4%; Pred. No. 1.1e-89;
Matches 263; Conservative 49; Mismatches 91; Indels 119; Gaps 11;
QY 5 PILGYWKIKGLVQPTLLLEYLPEEYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGD 64
Db |||||
QY 3 PILGYWKIKGLVQPTLLLEYLPEEYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGD 62
Db |||||
QY 65 VKLTQSMATIRYIADKHNHNGGCPKERAESMLEGAVLDIRYGVSRVIAYSKDFETLKVDF 124
Db |||||
QY 63 VKLTQSMATIRYIADKHNHNGGCPKERAESMLEGAVLDIRYGVSRVIAYSKDFETLKVDF 122
Db |||||
QY 125 LSKLPELAKMFEDRLCHKTYLNGDHVTHPDEFMLYDALDVLMDPCLDAPFKLVCFKKR 184
Db |||||
QY 123 LSKLPELAKMFEDRLCHKTYLNGDHVTHPDEFMLYDALDVLMDPCLDAPFKLVCFKKR 182
Db |||||
QY 185 IEAIPQIDKYLKSSKYIAWPLQGWQATGGDHPKSDHLPVPRGS----- 229
Db |||||
QY 183 IEAIPQIDKYLKSSKYIAWPLQGWQATGGDHPKSD- LYPGRSRRASVGSVHHHHHH 241
Db |||||
QY 230 -----MLRLKRNQIIS-----KRIVFTILK 250
Db ::||::||
QY 242 HGSVTKMSVYVVELLSFFYVTTETTFQKNRLFFRPSVMSKLSQSIGIRQHLKRVQLRELS 301
Db ::||::||
QY 251 QKAFKGNHPCVSVCTITYSRPHCLPDTLLKSLPMSSKTTLSMLPQVNVIGANSFSAETPV 310
Db ::||::||
QY 302 EAEVRQHRERAPALLT-SKLRFIPKPDGLRPIVM-----DYVVGARTFRREKRA 350
Db ::||::||
QY 311 D-----LKKENE-----TELANISGPHKKSTSTTKRARSKKK----- 345
Db ::||::||
QY 351 ERLTSRVKALFSVLNRYERARRPGLLGASVLGLDOTHRAWRTEFLRVRAQDPPPELYFKV 410
Db ::||::||
QY 346 -----ATDSV-SDKIDESVASY-----DSSTHLRRSRSRKKPNYNSSS 383
Db ::||::||
QY 411 DVTGAYDIPQDRUTEVTASIIKQNTVCVRRYAVVQKAAHGVHKKAFK-----HVST 464
Db ::||::||
QY 384 ESESEQISKATKKYKQKEEYEEVDEKSLKNSSSDEFE 425
Db ::||::||
QY 465 LFDLPYMRQFVAHLQETSPLRDAVWIEQSSSLNEASSGLFD 506
Db ::||::||

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 14:07:05 ; Search time 45.07 Seconds
(without alignments)
301.812 Million cell updates/sec

Title: US-09-724-296-4
Perfect score: 1799
Sequence: 1 DDHAPREMPDCLDKPIPWG.....KRLTARKRRSRKEVEDEK 371

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	360	18.2	320	1 YWJD_BACSU	P45864 bacillus su
2	111.5	5.6	517	1 YWJD_YEAST	P16140 saccharomyc
3	110	5.6	463	1 VAB2_DESSY	O06505 desulfuroco
4	108.5	5.5	492	1 VAB2_ARATH	P11574 arabidopsis
5	108	5.5	500	1 VAB2_CYACA	P48413 cyanidium c
6	104.5	5.3	470	1 VAB2_ARCFU	O29100 archaeoglob
7	104	5.3	463	1 VAB2_THESI	O32467 thermococu
8	104	5.3	490	1 VAB2_DROME	P31409 drosophila
9	103.5	5.2	2230	1 G0C4_HUMAN	Q13439 homo sapien
10	103	5.2	494	1 VAB2_HELVI	P31410 heliothis v
11	103	5.2	494	1 VAB2_WANSE	P31401 manduca sex
12	102	5.2	513	1 VAB1_BOVIN	P31407 bos taurus
13	101.5	5.1	453	1 VAB2_CHICK	P49712 gallus gall
14	101.5	5.1	465	1 VAB2_PYRAB	Q9uxu8 pyrococcus
15	101	5.1	491	1 VAB2_CAEEL	Q19626 caenorhabdi
16	101	5.1	800	1 CYAA_BACAN	P40136 bacillus an
17	100.5	5.1	497	1 PROD_MOUSE	Q9ux79 mus musculu
18	100.5	5.1	511	1 VAB2_HUMAN	P21281 homo sapien
19	100.5	5.1	511	1 VAB2_MOUSE	P50517 mus musculu
20	100	5.1	373	1 BIOF_AQUAE	O66875 aquifex aeo
21	100	5.1	513	1 VAB1_HUMAN	P15313 homo sapien
22	99.5	5.0	273	1 Y133_METJA	Q57597 methanococc
23	98.5	5.0	488	1 VAB1_GOSHI	Q43433 gossypium h
24	97.5	4.9	386	1 VAB2_GOSHI	P31408 bos taurus
25	97.5	4.9	511	1 VAB2_BOVIN	O57729 pyrococcus
26	96.5	4.9	465	1 VAB2_PYRHO	P22550 candida tro
27	96	4.9	511	1 VAB2_CANTR	P72241 providencia
28	95	4.8	365	1 GCPE_PROST	O66431 dudbe virus
29	95	4.8	4036	1 RRPL_DUGBV	Q38680 acetabulari
30	94.5	4.8	492	1 VAB2_ACEAT	Q38681 acetabulari
31	93.5	4.7	492	1 VAB1_MOUSE	P28660 mus musculu
32	93.5	4.7	648	1 NCPI_MOUSE	P55161 rattus norv
33	93.5	4.7	1128	1 NCPI_RAT	

ALIGNMENTS

RESULT	ID	YWJD_BACSU	STANDARD	PRT	320 AA
AC	P45864	1			
DT	01-NOV-1995	(Rel. 32, Created)			
DT	01-NOV-1995	(Rel. 32, Last sequence update)			
DT	20-AUG-2001	(Rel. 40, Last annotation update)			
DE	HYPOTHETICAL 36.9 KDA PROTEIN IN ACDA 5'REGION.				
GN	YWJD				
OS	Bacillus subtilis.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;				
OC	Bacillus/Staphylococcus group; Bacillus.				
OX	NCBI_TaxID=1423;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=168;				
RA	Glaser P., de la Fuente V., Danchin A.;				
RL	Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	EMBL; Z49782; CAA89865.1; -				
DR	EMBL; Z49782; CAB15748.1; -				
DR	Subtilist; BGI1309; YWJD.				
DR	Hypothetical protein; Complete proteome.				
KW	SEQUENCE 320 AA; 36896 MW; 2FD35A710648E002 CRC64;				
SQ					
Query Match	18.2%;	Score 360;	DB 1;	Length 320;	
Best Local Similarity	29.8%;	Pred. No. 4.7e-21;			
Matches	84;	Conservative	61;	Mismatches 109;	Indels 28; Gaps 8;
QY	48	TIORDGLESVKQLGTQNVLDLIKLVNHNHFGIHFMRVSSDLFFPPASH--AKYGYTLEFA	105		
DB	34	TERKEALLTVKANLRNTM---RTLHYTIIGHIPLRYESSIVPLATHPDVMDVFTPPF-	89		
QY	106	QSHLEEVGKLANKYNHRLTMIPGQYQTATSPREVVDVAIRDLAYHDEILSRMKLNEDLN	165		
DB	90	QKEFREIGELVKTHQRTSFHPNQFTLFTSPKESVTKNAVTDMAVHYRMLEAMGIAADR--	147		
QY	166	KDAVLI IHGGTFECKKETLDRFKRYORLSDSVKARVLNDVDSWSVDLLPLCCQELN	225		
DB	148	---SVINIHGGAYGNKDTATAGFHONIKQLQOEIKERTWLENDKTYTTEETLQVCEQED	205		
QY	226	IPLVLDDWHHHNIVPGTLREGSLDLMLPIPTIRETWKRGITQKQHYSESADPTAISGMKR	285		
DB	206	VFFVDFDHFHEYANP---DDHADLNVALPRMKTWERIGLQPKVHLSSPKSRQAI-----	256		
QY	286	RAHSRDRV---FDFPPCD-----PTMDLMIEAKEKEQAVFEL	318		

Db 257 RSHADYVDANFLPLLERFRQWGTNTDFMIEAKQKQKALLRL 298

RESULT 2

VATB_YEAST STANDARD; PRT; 517 AA.

AC P16140; P32123;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-OCT-1994 (Rel. 30, Last annotation update)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

DE (VACUOLAR ATP SYNTHASE SUBUNIT B (EC 3.6.1.34) (V-ATPASE B SUBUNIT))

DE (VACUOLAR PROTON PUMP B SUBUNIT) (V-ATPASE 57 KDA SUBUNIT).

GN VMA2 OR VAT2 OR YBR137C OR YBR1002.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes.

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=89109197; PubMed=2521486;

RA Nelson H., Mandiyan S., Nelson N.;

RA "A conserved gene encoding the 57-kDa subunit of the yeast vacuolar H⁺-ATPase.";

RL J. Biol. Chem. 264:1775-1778(1989).

RN [2]

RP REVISIONS.

RA Nelson H., Mandiyan S., Nelson N.;

RL J. Biol. Chem. 264:5313-5313(1989).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=S288C;

RX MEDLINE=94378717; PubMed=8091856;

RA Becam A.-M., Cullin C., Grzybowski E., Lacroix F., Nasr F.,

RA Ozier-Kalogeropoulos O., Palucha A., Slonimski P.P., Zagulski M.,

RA Herbert C.J.;

RT "The sequence of 29.7 kb from the right arm of chromosome II reveals 13 complete open reading frames, of which ten correspond to new genes.";

RL Yeast 10:S1-S11(1994).

RN [4]

RP SEQUENCE OF 103-517 FROM N.A.

RX MEDLINE=92156101; PubMed=1371275;

RA Puopolo K., Kumamoto T., Adachi I., Mager R., Forgac M.;

RT "Differential expression of the 'B' subunit of the vacuolar H⁺-ATPase in bovine tissues.";

RL J. Biol. Chem. 267:3696-3706(1992).

CC -!- FUNCTION: NONCATALYTIC SUBUNIT OF THE PERIPHERAL V1 COMPLEX OF VACUOLAR ATPASE. V-ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS. IT IS AN ELECTROGENIC PROTON PUMP THAT GENERATES A PROTON MOTIVE FORCE OF 180 MV, INSIDE POSITIVE AND ACIDIC, IN THE VACUOLAR MEMBRANE VESICLES.

CC -!- SUBUNIT: V-ATPASE IS AN HETEROMULTIMERIC ENZYME COMPOSED OF A PERIPHERAL CATALYTIC V1 COMPLEX (MAIN COMPONENTS: SUBUNITS A, B, C, D, E, AND F) ATTACHED TO AN INTEGRAL MEMBRANE V0 PROTON PORE COMPLEX (MAIN COMPONENT: THE PROTEOLIPID PROTEIN).

CC -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.

CC -!- CAUTION: REF.4 SEQUENCE WAS INCORRECTLY THOUGHT TO ORIGINATE FROM BOVINE.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: J04450; AAA6890.1; -

EMBL: X75891; CAAS3486.1; -

EMBL: Z35996; CAAS5084.1; -

EMBL: M8130; AAA30389.1; -

PIR: S45996; S45996.

DR PIR: B42254; B42254.

DR PIR: S46570; S46570.

DR YRPD; R560; -

DR SGD; S0000331; VMA2.

DR InterPro; IPR003255; ATP_synt_B_c.

DR InterPro; IPR000194; ATPase_alpha_beta.

DR Pfam; PF00006; ATP_synt_ab; 1.

DR Pfam; PF01038; ATP_synt_B_c; 1.

DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.

KW Hydroxylase; ATP synthetase; Hydrogen ion transport.

FT CONFLICT 79 79 V -> L (IN REF. 1).

FT CONFLICT 227 227 Q -> R (IN REF. 4).

FT CONFLICT 500 517 DTRSGKKKSDASOESLI -> TQAPVRRRTPAKKNL (IN REF. 1).

FT SEQUENCE 517 AA; 57749 MW; 02A7DEC571EFF7C2 CRC64;

Query Match 5.6%; Score 111.5; DB 1: Length 517;

Best Local Similarity 20.0%; Pred. No. 0.26;

Matches 82; Conservative 62; Mismatches 140; Indels 125; Gaps 18;

QY 27 LNTILRSKKEVPF-----CSRTCRITTTIQDGLSVKQLGTQNVLDLIKLVENH 76

Db 161 MNSIARGQKIPFSASGLPHNETAAQICRQAGLVRP-----TKDVHD-----GHEE 206

QY 77 NFGIHFMRVSSDLFPFASHAKYGYTLEFAQSHLEVG-----KLANK-----YNH 121

Db 207 NFSIVFAAMGNL-----ETARFPKQDFEENGSLERTSLFLNLANDPRIIITP 256

QY 122 RLTHMPGQYQTIASPREVVV-----DSAIRDL-AYHDEILSRMKLNQLNKDAVLII 172

Db 257 RLALTATAYLAYQTERHVLTLTDMSSYADALREVSAAREVEVPGRRGYPCGYMTDLSTIV 316

QY 173 HUGTEPECKKETLDRFRKNYQRLSDSVKARLVLENDVSVQDLPLPLCOELNIPVLVOW 232

Db 317 ERAGRVEGRNGSIQT-----PLITMPNDITHTIPDLTGYITEGIFVDROL 364

QY 233 HHNIIVPTLREGSLDMLPIPTIRETWTRKGTOKOH-----YS---ESADPTA 279

Db 365 HNKGIYP-----PINVLPSSLRLMKSALGEGMTRKDHGVDVSNQLYAKYAGKDAAMKA 418

QY 280 ISGMKRRRAHSDRVDFPPCDPTMDLMEAKEK-----EQAVFE-----LCR 320

Db 419 VVGEELSLIEDK-----LSLEFLKFEKFTITQGYEDRTVFESLDQAWSLLR 466

QY 321 RY--ELQNFPCPLEIMGPEYDQTRD--GYPPGAERKRLTARKRRSRKEE 365

Db 467 IYPKEMLNRTSP-KILDEFYDRADDDEDEDPDTRSSGKKKSDASQEE 514

RESULT 3

VATB_DESSY STANDARD; PRT; 463 AA.

AC 006505;

DT 20-AUG-2001 (Rel. 40, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE V-TYPE ATP SYNTHASE BETA CHAIN (EC 3.6.1.34) (V-TYPE ATPASE SUBUNIT B).

DE H).

GN ATPB.

OS Desulfurococcus sp. (strain SY).

OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;

OC Desulfurococcus.

OX NCBI_TaxID=59822;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=97320421; PubMed=9177272;

RA Shibui H., Hamamoto T., Yoshida M., Kagawa Y.;

RT "The stabilizing residues and the functional domains in the hyperthermophilic V-ATPase of Desulfurococcus".

RL Biochem. Biophys. Res. Commun. 234:341-345(1997).

CC -!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON GRADIENT ACROSS THE MEMBRANE. THE ARCHAEL BETA CHAIN IS A


```

CC      REGULATORY SUBUNIT.
CC      -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U96487; AAB64417.1; ALT INIT.
DR      InterPro: IPR003255; ATP_synt_B_c.
DR      InterPro: IPR000194; ATPase_alpha_beta.
DR      Pfam: PF00006; ATP_synt_ab; 1.
DR      Pfam: PF01038; ATP_synt_B_c; 1.
DR      PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
DR      Hydrolase; ATP synthesis; Hydrogen ion transport.
DR      KW HYDROLASE; ATP SYNTHESIS; HYDROGEN ION TRANSPORT.
DR      SQ SEQUENCE 463 AA; 52051 MW; 8D94195D02D50D9A CRC64;

Query Match          5.6%; Score 110; DB 1; Length 463;
Best Local Similarity 19.1%; Pred. No. 0.29;
Matches 71; Conservative 55; Mismatches 153; Indels 92; Gaps 13;

QY 27 LNTILSRMKERVPCSRCTRTTIQDGLSVKQLGTONVLDLKLVEWNHNFQIHFMVRS 86
DB 139 MNTLVRGQKLPFGSGLPNNMLAAQIAQKAVLGEE-----EQFAVVF----- 182
QY 87 SDLFPFASHAKYGYTLE---FAQSHLEEVCKLANK-----YHRLTMTHPG 128
DB 183 -----AAMGITYEAEFNFKKSFEEGATRAVFLFLNLADDPALRIITPRMALTYA 233
QY 129 QYTIQIASPREVVY-----DSAIRDL-AYHDEILSRMKLNEQLNKDAVLIHLGGTPE 179
DB 234 EYLAFDYDMQVILVILDMTNYAALREISAAREEVPGRGYPGYMTDLATIVERGGRV 293
QY 180 GKKETLDRFRKNYQRLSDSVKARVLENDVSVSDVODLLPLCOELNIPVLWDDHHNIVP 239
DB 294 GKKGSIOM-----PILTMPDDDDITHPDLTGYTEGOVLVSRELHRKGIYTP 341
QY 240 GTLREGSLDMLPIPTIRETWTRKGIKQKH-----YSES---ADPTAISGMKRR 286
DB 342 -----PIDVLPSSLRLMKDGIKGRTRDHPQLAQOLVAAVAGRSRLDLVAVVGEAL 395
QY 287 AHSRVDVDFPPCDPTMDLMEAKEQAVFE-LCRYEYELQNPPCPLEIMGPEVDQTRGY 345
DB 396 SETDRKYLKFADRFEREFTAQRYDEDRSIFETLDLGWELL-----AELPESELKRVKRY 450
QY 346 ---YPPGAEKRR 353
DB 451 ILKYHPKRYKR 461

RESULT 4
VATB_ARATH
ID VATB_ARATH STANDARD; PRT; 492 AA.
AC P11574; Q42205;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE VACUOLAR ATP SYNTHASE SUBUNIT B (EC 3.6.1.34) (V-ATPASE B SUBUNIT)
DE (VACUOLAR PROTON PUMP B SUBUNIT) (V-ATPASE 57 KDA SUBUNIT).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=890533962; PubMed=2903860;
RA Manolison M.F., Ouellette B.F.F., Fillion M., Poole R.J.;
RT "cDNA sequence and homologies of the '57-kDa' nucleotide-binding

```

```

RT      subunit of the vacuolar ATPase from Arabidopsis.";
RL      J. Biol. Chem. 263:17987-17994(1988).
RL      [2]
RP      SEQUENCE OF 241-340 FROM N.A.
RC STRAIN=CV; COLUMBIA; TISSUE=Seedling;
RA Desprez T., Anselme J., Chiapello H., Caboche M., Hofte H.;
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
CC      -!- FUNCTION: NONCATALYTIC SUBUNIT OF THE PERIPHERAL V1 COMPLEX OF
CC      VACUOLAR ATPASE. V-ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY
CC      OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS.
CC      -!- SUBUNIT: V-ATPASE IS AN HETEROMULTIMERIC ENZYME COMPOSED OF A
CC      PERIPHERAL CATALYTIC V1 COMPLEX (MAIN COMPONENTS: SUBUNITS A, B,
CC      C, D, E, AND F) ATTACHED TO AN INTEGRAL MEMBRANE V0 PROTON PORE
CC      COMPLEX (MAIN COMPONENT: THE PROTEOLIPID PROTEIN).
CC      -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; J04185; AAC36485.1; -.
DR      EMBL; Z29126; CAAB2390.1; -.
DR      PIR; A31886; A31886.
DR      InterPro: IPR003255; ATP_synt_B_c.
DR      InterPro: IPR000194; ATPase_alpha_beta.
DR      Pfam: PF00006; ATP_synt_ab; 1.
DR      Pfam: PF01038; ATP_synt_B_c; 1.
DR      PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
DR      Hydrolase; ATP synthesis; Hydrogen ion transport.
DR      KW CONFLICT 303 303 R -> V (IN REF. 2).
DR      SQ SEQUENCE 492 AA; 54739 MW; 2BF630AA79EEAC7C CRC64;

Query Match          5.5%; Score 108.5; DB 1; Length 492;
Best Local Similarity 20.0%; Pred. No. 0.41;
Matches 69; Conservative 60; Mismatches 123; Indels 93; Gaps 16;

QY 27 LNTILSRMKERV-----CSRTCRITTIQDGLSVKQLGTONVLDLKLVEWNH 76
DB 158 MNSIARGQKIPLSAAGLPHEIAAICR-----QAGL--VKRL--EKTVDLLE---DH 204
QY 77 ---NFGIHFMRVSSDLFPFASHAKYGYTLFEFAQSHLEEVG-----KLANK----- 118
DB 205 SEDNFAIVFAANGVNM-----ETAQPFKRDFEENGSMERYTLFLNLANDPTIERI 254
QY 119 YNHLRTMHPGOVYTIASPREVVV-----DSAIRDL-AYHDEILSRMKLNEQLNKDAV 169
DB 255 ITPRIALTAEYLAECGKHVILVITDMSSYADALREYSAAREEVPGRGYPGYMYTDLA 314
QY 170 LTIHLGGTFEGKKETLDRFRKNYQRLSDSVKARVLENDVSVSDVODLLPLCOELNIPLV 229
DB 315 TIYERAGRIEGRKGSITQI-----PILTMPDDDDITHPDLTGYTEGOVLV 362
QY 230 LDWHHNIVPGTLREGSLDMLPIPTIRETWTRKGIKQKH-----YSESADPTAISG 282
DB 363 RQLHNRQIYIP-----PINVLPSSLRLMKSAIGCMTRKDSVSNQLYANYATKGVQA 416
QY 283 MKRRAHSRVRVDFPPCDPTMDLM-LTEAKEQAVFELCRRYELQN 326
DB 417 MKAVVGEAL-----SSEDLILYLFLELDKPERKFMQCAVYDTRN 454

RESULT 5
VATB_CYACA
ID VATB_CYACA STANDARD; PRT; 500 AA.
AC P48413;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)

```

DE VACUOLAR ATP SYNTHASE SUBUNIT B (EC 3.6.1.34) (V-ATPASE B SUBUNIT)
DE (VACUOLAR PROTON PUMP B SUBUNIT)
OS Cyanidium caldarium.
OC Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;
OC Cyanidium.
OX NCBI_TaxID=2771;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95345110; PubMed=7619836;
RA Ziegler K., Hauska G., Nelson N.;
RT "Cyanidium caldarium genes encoding subunits A and B of V-ATPase.";
RL Biochim. Biophys. Acta 1230:202-206(1995).
CC -!- FUNCTION: NONCATALYTIC SUBUNIT OF THE PERIPHERAL V1 COMPLEX OF
CC VACUOLAR ATPASE. V-ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY
CC OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS.
CC -!- SUBUNIT: V-ATPASE IS AN HETEROMULTIMERIC ENZYME COMPOSED OF A
CC PERIPHERAL CATALYTIC V1 COMPLEX (MAIN COMPONENTS: SUBUNITS A, B,
CC C, D, E, AND F) ATTACHED TO AN INTEGRAL MEMBRANE V0 PROTON PORE
CC COMPLEX (MAIN COMPONENT: THE PROTEOLIPID PROTEIN).
CC -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U17101; AAA85821.1; -
DR InterPro: IPR003255; ATP_synt_B_C.
DR InterPro: IPR000194; ATPase_alpha_beta.
DR Pfam: PF00006; ATP_synt_ab; 1.
DR Pfam: PF01038; ATP_synt_B_C; 1.
DR PROSITE: PS00152; ATPASE ALPHA BETA; 1.
KW Hydrolase; ATP synthesis; Hydrogen ion transport.
SQ SEQUENCE 500 AA; 56433 MW; 4A882382E8E9CF5 CRC64;

Query Match 5.5%; Score 108; DB 1; Length 500;
Best Local Similarity 18.7%; Pred. No. 0.46;
Matches 66; Conservative 52; Mismatches 129; Indels 106; Gaps 12;

Qy 5 PREMFCDLCKIPWRGLGYACILNTILSRMKERYF-----CSRTRIT-- 48
Db 142 PEEMFE-----TGSSIDVMNSIARGQKIPFLSGAGLPNEVAQAICQVCLVSTCT 193
Qy 49 -IQRDGLSEVKQLGTQNVLDLKLVEVNHNPFIHFMVSSDLPPFASHAKYGYTLEFAQS 107
Db 194 LVKRSKGD-----EEDFAIVFAAMGVNM-----ETARFFRQ 224
Qy 108 HLEEVG-----KLANK-----YNHRLTHPGGYTOIASPREVV-----DSA 144
Db 225 DFEENGAMERTVTLFLNLANPTIERITPRALTAFAEYLAYEKGKHLVLTIDMSAYADA 284
Qy 145 IRLD-LAYHDEILSRMKNEQLNDAVLIHLGGTFEKGKKTLDLRFKKNYQRLSDSVKARL 203
Db 285 LREVSAAEEVPGRRGYPGYMTDLATYIYERAGVRGPGSIQL-----PIL 332
Qy 204 VLENDVSVSWQDILLPLCOELNIPVLVDWHHNNIPVPTLREGSLDMLPLIPTIRETWTR 262
Db 333 TMPNDITHTPIPDLTGYITEQIVLSRELHAKGIYP-----PINVLPSLSRLMKSAIGE 386
Qy 264 GITOKOHYESADPTALSGMKRRAHSRDVDFPPCDPTMDLMTAEAKEQAVF 316
Db 387 GMTKRKSDSVSNQLYAAYAMGKDALAMRAV-----VGVEALSQEDLLY 429

RESULT 6
VATB_ARCFU
ID VATB_ARCFU STANDARD; PRT; 470 AA.
AC Q29100;
DT 20-AUG-2001 (Rel. 40, Created)

DT 20-AUG-2001 (rel. 40, Last sequence update)
DT 20-AUG-2001 (rel. 40, Last annotation update)
DE V-TYPE ATP SYNTHASE BETA CHAIN (EC 3.6.1.34) (V-TYPE ATPASE SUBUNIT
DE B).
GN ATPB OR AF1167.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriplides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Cocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC GRADIENT ACROSS THE MEMBRANE. THE ARCHAEL BETA CHAIN IS A
CC REGULATORY SUBUNIT.
CC -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF001023; AAB90073.1; ALT_INIT.
DR TIGR: AF1167; -
DR InterPro: IPR003255; ATP_synt_B_C.
DR InterPro: IPR000194; ATPase_alpha_beta.
DR Pfam: PF00006; ATP_synt_ab; 1.
DR Pfam: PF01038; ATP_synt_B_C; 1.
DR PROSITE: PS00152; ATPASE ALPHA BETA; 1.
KW Hydrolase; ATP synthesis; Hydrogen ion transport; Complete proteome.
SQ SEQUENCE 470 AA; 52874 MW; C57B2F845FB873E2 CRC64;

Query Match 5.3%; Score 104.5; DB 1; Length 470;
Best Local Similarity 20.2%; Pred. No. 0.79;
Matches 51; Conservative 45; Mismatches 93; Indels 63; Gaps 9;

Qy 144 ATRDL-AYHDEILSRMKNEQLNDAVLIHLGGTFEKGKKTLDLRFKKNYQRLSDSVKAR 202
Db 254 ALREISAAEEVPGRRGYPGYMTDLATYIYERAGVRGPGSIQL-----PI 301
Qy 203 VLENDVSVSWQDILLPLCOELNIPVLVDWHHNNIPVPTLREGSLDMLPLIPTIRETWTR 262
Db 302 LIMPDDITHTPIPDLTGYITEQIVLSRELHAKGIYP-----PINVLPSLSRLMKEGIG 355
Qy 263 KQITQKH-----YESAD-----PTAISGMKRRRAHSRDVDFPPCDPTMDLMTAEAK 309
Db 356 EGYTREDHPQWNDQWYAAYAEQVLDRLGLVAIVGEALSERDLF-----LKFA 403
Qy 310 EKEQAVFELCRRYELONPPCPLEI-----MCPEVDQTR-----DGYVPCAEKRLTAR 357
Db 404 DEFERRFVQGGRYEDRDIEYLDLGNELLAMLPERELTKVERKFTKYHP-----KYR 456
Qy 358 KRRSRKEVEED 369
| : : : | | |

Db 457 KKEAPPEATEE 468

RESULT 7

```
VATB_THESI
ID VATB_THESI STANDARD; PRT: 463 AA.
AC O32467;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE V-TYPE ATP SYNTHASE BETA CHAIN (EC 3.6.1.34) (V-TYPE ATPASE SUBUNIT
DE B).
DE ATPB.
GN Thermococcus sp. (strain KI).
OS Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Thermococcus.
OC NCBI_TaxID=35749;
OX [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=98036007; PubMed=9370240;
RX Iida T., Kanai S., Inatomi K., Kamagata Y., Maruyama T.;
RT "Alpha- and beta-subunits of a V-type membrane ATPase in a
RT hyperthermophilic sulfur-dependent archaeum, Thermococcus sp. KI.";
RL Biochim. Biophys. Acta 1329:12-17(1997).
CC -1- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC GRADIENT ACROSS THE MEMBRANE. THE ARCHAEAL BETA CHAIN IS A
CC REGULATORY SUBUNIT.
CC -1- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D88772; BAA23343.1;
DR InterPro: IPR003255; ATP_synt_b_c.
DR InterPro: IPR000194; ATPase_alpha_beta.
DR Pfam: PF00006; ATP_synt_ab; 1.
DR Pfam: PF01038; ATP_synt_b_c; 1.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
KW Hydrolase; ATP synthetase; Hydrogen ion transport.
SQ SEQUENCE 463 AA; 52168 MW; 2F0EC25DA60462DF CRC64;
```

Query Match 5.38; Score 104; DB 1; Length 463;
Best Local Similarity 18.98; Pred. No. 0.85;
Matches 70; Conservative 55; Mismatches 154; Indels 92; Gaps 13;

```
QY 27 LNTILSRKERVFCSTCRITTIQDGLSVKQLGTQNVLDLIKLVENHNFGIHFMRVS 86
DB 139 MNTLVGQKLPFGSGSLPHNMTAAQTARQAKVLGEE-----EQFAVVF---- 182
QY 87 SDLFPFASHAKYGYTLE---FAQSHLEEVGKLANK-----YNHRLTMHPG 128
DB 193 -----AANGITYEANFFKSPETGAERAVLFNLADDPATERIITPRMALTVA 233
QY 129 QYTIASPREVV-----DSAIRDL-AVHDEILSRMKLNEOLNKAVALIHLGTFE 179
DB 234 EYLFADYDMQVLVILDTMTNVAEALREISAAREVPGRRGPGMYMTDLATIVERAGRVR 293
QY 180 GKKTLDLFRNRYORLSDSVKARVLENDVSVSQDLLPLCQELNIPFLVDLWHHNIIVP 239
DB 294 GKGSITQM-----PIITMPDDDDITHPIPLDTGTGTVGQIVLSRELHKKGIYP 341
QY 240 GTLREGSLDLPLIPTIRETWRGTGKQKOH-----YSES-----ADPTATSGMKRR 286
DB 342 -----PIDVLPSLSRLMKDGIKGRTRFEDIPQLSQLYAAYAGRSRLDLVAVVGGEAL 395
QY 287 AHSRDFVDFPPCDPTMDLMIEAKEQAVFE-LCRRYELQNPCCPLEIMGPEDQTRDGY 345
DB 396 SETDRKYLKFAFRFREFVAQRYEDRSIPETLDLGWELL-----AELPESELKRVKEY 450
```

QY 346 ---YPPGAEKR 353

Db 451 ILKYHPKYRKR 461

RESULT 8

```
VATB_DROME
ID VATB_DROME STANDARD; PRT: 490 AA.
AC P31409; Q9VG52;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE VACUOLAR ATP SYNTHASE SUBUNIT B (EC 3.6.1.34) (V-ATPASE B SUBUNIT)
DE (VACUOLAR PROTON PUMP B SUBUNIT) (V-ATPASE 57 KDA SUBUNIT).
GN VHA55 OR CG17369.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=EYA; TISSUE=Head;
RX MEDLINE=97094810; PubMed=8940044;
RA Davies S.A., Kelly D.C., Goodwin S.F., Yang S., Sozen M.A.;
RA Kaiser K., Dow J.A.T.;
RT "Analysis and inactivation of vha55, the gene encoding the vacuolar
RT ATPase B-subunit in Drosophila melanogaster reveals a larval lethal
RT phenotype.";
RL J. Biol. Chem. 271:30677-30684(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H.J., Adayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Besley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.F., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Correll J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laspo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
```



```
QY 31 LRSKRVFCSRTCTRTTTIORDGLESV-----KOL-----GTQNVLDLIKLV 73
DB 1622 LKALEDRLSESAAKLAEUKRAEKQIAIAIKKOLLSQMEKEEQYKKGTFESHLSEINTKL 1681
QY 74 WHNFIHFMW-----VSSD-----LFPASHAKYGYTLEFAQSHLEEVGKLANKYNHRL 123
DB 1682 QEREREVHILEEKLKSVSSOSETLIVPSAKNVAAT---EQEADSGCGVQKTYEKKI 1738
QY 124 THHPGYGTO-----IASPREVVVDSAIRDLAYHDEILSRMKLNE-QLNKDAVLIIHLG 175
DB 1739 SVLQRLNTEKEKLLQROGQEKETVSSHFMRCQYQERLIKLEHAEAKOHEQDSMIGHL- 1797
QY 176 GFEGKKTLDLFRKKNYRL-----SDSVKARLVLEN--DDVSVSVDLLPLQO- 222
DB 1798 -----QEELEEKKNYSILIVAOHVEKEGKNIAQONKQNVFDDVQKTLQEKELTQOI 1851
QY 223 -----ELNPLVLDWHHNIHVPGLT-----EGSLDMLPLIPTRETWTTRKGTQKHVSE 273
DB 1852 LEQIKELDSCLVROKEVHRVEMBELTSKYKLOAQOQDGRNKPTELLEENTEESKSH 1911
QY 274 SADPTAISGMKRAHSDRVDFPPCDPTWDLMTAEAKEQAVFELCR-----RYELQ 325
DB 1912 LVQPKLLSNMEAQ--HNDLEFKLAGAE-----REKOKLGEIVRLQKDLRLMRKEHQ 1961
QY 326 NPPCPLEIMGPEYDQTRDGYPPGAEKRLTARKRRSRKEEVE 367
DB 1962 Q-----ELEILKKEYDQERE-----EKIKOEQEDLE 1987

RESULT 10
VATB_HELVI
ID VATB_HELVI STANDARD; PRT: 494 AA.
AC P31410;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE VACUOLAR ATP SYNTHASE SUBUNIT B (EC 3.6.1.34) (V-ATPASE B SUBUNIT)
DE (VACUOLAR PROTON PUMP B SUBUNIT).
GN VHA55.
OS Heliothis virescens (Noctuid moth) (Owlet moth).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Noctuoidea; Noctuidae; Heliothinae; Heliothis.
OX NCBI_TaxID=7102;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Midgut;
RX MEDLINE=92027804; PubMed=1834020;
RA Gill S.S., Ross L.S.;
RT "Molecular cloning and characterization of the B subunit of a
RT vacuolar H(+)-ATPase from the midgut and Malpighian tubules of
RT Helicoverpa virescens."
RL Arch. Biochem. Biophys. 291:92-99(1991).
CC -!- FUNCTION: NONCATALYTIC SUBUNIT OF THE PERIPHERAL V1 COMPLEX OF
CC VACUOLAR ATPASE. V-ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY
CC OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS.
CC -!- SUBUNIT: V-ATPASE IS AN HETEROMULTIMERIC ENZYME COMPOSED OF A
CC PERIPHERAL CATALYTIC V1 COMPLEX (MAIN COMPONENTS: SUBUNITS A, B,
CC C, D, E, AND F) ATTACHED TO AN INTEGRAL MEMBRANE V0 PROTON PORE
CC COMPLEX (MAIN COMPONENT: THE PROTEOLIPID PROTEIN).
CC -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR InterPro: IPR003255; ATP_synt_B_C.
DR InterPro: IPR000194; ATPase_alpha_beta.
DR Pfam: PF00006; ATP_synt_ab; 1.
DR Pfam: PF01038; ATP_synt_B_C; 1.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; 1.
KW Hydrolase; ATP synthetase; Hydrogen ion transport.
SQ SEQUENCE 494 AA; 54893 MW; 2D20C581D8AE4C76 CRC64;

Query Match 5.2%; Score 103; DB 1; Length 494;
Best Local Similarity 19.7%; Pred. No. 1.1;
Matches 61; Conservative 48; Mismatches 114; Indels 86; Gaps 13;

QY 7 EMFCLDKPI-PW-----RGLGYACILNLSMKERVF-----CSRTCR 45
DB 132 DFLDIQOGIPNWSKYIPEEMIQTCISDAIDVNSIARQKIPISAGLPHNEIAAQICR 191
QY 46 ITTIORDGLESVKOLGTQNVLDLIKLVENHNFHGMKRVSSDLFPFASHAKYGYTLEFA 105
DB 192 -----QAGL--VKIPG-KSVLD-----DHEDNFAIVFAAMGVNM-----ETARFF 228
QY 106 QSHLEEVCKLANK-----YNHRLTMHPCQYTCIASPREVVV-----D 142
DB 229 KODFEENGSMENVCLFLNLANDPTRITRPLALTAAEFLAYQCEKHVILVLTDMSSYA 288
QY 143 SAIRDL-AVHDEILSRMKLNEQLNKDAVLIIHLGTFEGKKTLDLDRKKNYQRLSDSVKA 201
DB 289 EALREVSAAREVPORGFPGMYTDLATIERAGRVGECNCSITQI-----P 336
QY 202 RLVLNDDVSVSVODLLPLCOELNPLVLDWHHNIHVPGLTREGSLDMLPLIPTRETWT 261
DB 337 ILTMPNDITTHIPDLTGIVTEGQIYVDRQLHNRQIYP-----PVNLPVLSRLMSAI 390
QY 262 RKGITQKOH 270
DB 391 GEGMTRKDH 399

RESULT 11
VATB_MANSE
ID VATB_MANSE STANDARD; PRT: 494 AA.
AC P31401;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE VACUOLAR ATP SYNTHASE SUBUNIT B (EC 3.6.1.34) (V-ATPASE B SUBUNIT)
DE (VACUOLAR PROTON PUMP B SUBUNIT).
GN VHA55.
OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Sphingioidea; Sphingidae; Sphinginae; Manduca.
OX NCBI_TaxID=7130;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Midgut;
RX MEDLINE=92379093; PubMed=1387326;
RA Novak F.J., Graf R., Waring R.B., Wolfersberger M.G., Wieczorek H.,
RA Harvey W.R.;
RT "Primary structure of V-ATPase subunit B from Manduca sexta midgut."
RT Biochim. Biophys. Acta 1132:67-71(1992).
CC -!- FUNCTION: VACUOLAR ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY
CC OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS. THE B SUBUNIT
CC IS NONCATALYTIC BUT COMBINES WITH OTHER SUBUNITS TO FORM THE
CC CATALYTIC COMPLEX. V-ATPASE IS RESPONSIBLE FOR ENERGIZING
CC ELECTROPHORETIC K+/2H+ ANTIPORT BY GENERATING A TRANSMEMBRANE
CC VOLTAGE OF MORE THAN 200MV.
CC -!- SUBUNIT: V-ATPASE IS AN HETEROMULTIMERIC ENZYME COMPOSED OF A
CC PERIPHERAL CATALYTIC V1 COMPLEX (MAIN COMPONENTS: SUBUNITS A, B,
CC C, D, E, AND F) ATTACHED TO AN INTEGRAL MEMBRANE V0 PROTON PORE
CC COMPLEX (MAIN COMPONENT: THE PROTEOLIPID PROTEIN).
CC -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC
```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X64354; CAA45706.1; -
 DR PIR; S19206; S19206.
 DR PIR; S24387; S24387.
 DR InterPro; IPR003255; ATP_synt_B_C.
 DR InterPro; IPR000194; ATPase_alpha_beta.
 DR Pfam; PF00006; ATP_synt_ab; 1.
 DR Pfam; PF01038; ATP_synt_B_C; 1.
 DR PROSITE; PS00152; ATPase_ALPHA_BETA; 1.
 KW Hydrolase; ATP synthesis; Hydrogen ion transport.
 SQ SEQUENCE 494 AA; 54904 MW; 8A8A3A2C04A9CF06 CRC64;

Query Match 5.2%; Score 103; DB 1; Length 494;
 Best Local Similarity 19.7%; Pred. No. 1.1;
 Matches 61; Conservative 48; Mismatches 114; Indels 86; Gaps 13;

QY 7 EMFCDLKPI-PW-----RGLGYACLTILSRMKERVF-----CSRTCR 45
 DB 132 DFLDLOGQPINPWSRIYFEEMIQGISAIQVNSIARQKIPISAGLPHNEIAAQICR 191
 QY 46 ITTIQDGLSVKQIGTQNVLDLKLVEWNNHFGIHFMRVSSDLFPFASHAKYGYTLEFA 105
 DB 192 -----QAGL--VKIPG-KSVLD-----DHEDNFAIVFAAMGVNM-----ETARFF 228
 QY 106 OSHLEEVCKLANK-----YNRLTMHFGQYTOIASPREVV-----D 142
 DB 229 KQDFEENGSMENVLFLMLANDPTIERIITPRALTAETAEFLAYQCEKHVILVILTDMSSYA 288
 QY 143 SAIRDL-AYHDEILSRMKLNQKDAVLIILHGGTPEGKKTLDLDRFKNYQRLSDSVKA 201
 DB 289 EALREVSAAAREVPGRRGFGPYMTDLATIVERAGRVEGRNGSITQI-----P 336
 QY 202 RLVLNDDSVSVODLLPLCOELNIPVLVDWHHNIIVPGTLREGSLDMLPIPTIRTW 261
 DB 337 ILTMPNDIITHIPDLTGITYEGQIYVDROLHNRQIYP-----PNNVLPSPSLRMSAI 390
 QY 262 RKGITQKOH 270
 DB 391 GEGMTRKDH 399

RESULT 12
 VABI_BOVIN STANDARD; PRT; 513 AA.
 AC P31407;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE VACUOLAR ATP SYNTHASE SUBUNIT B, KIDNEY ISOFORM (EC 3.6.1.34) (V-
 DE ATPASE B1 SUBUNIT) (VACUOLAR PROTON PUMP B ISOFORM 1) (ENDOMEMBRANE
 DE PROTON PUMP 58 KDA SUBUNIT).
 GN ATP6B1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 504-513.
 RC TISSUE=Kidney;
 EX MEDLINE=92226822; PubMed=1373501;
 RA Nelson R.D., Guo X.-L., Masood K., Brown D., Kalkbrenner M.,
 RA Gluck S.;
 RT "Selectively amplified expression of an isoform of the vacuolar H(+)-
 RT ATPase 56-kilodalton subunit in renal intercalated cells.";

RL Proc. Natl. Acad. Sci. U.S.A. 89:3541-3545(1992).
 CC -!- FUNCTION: NON-CATALYTIC SUBUNIT OF THE PERIPHERAL V1 COMPLEX OF
 CC VACUOLAR ATPASE. V-ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY
 CC OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS.
 CC -!- SUBUNIT: V-ATPASE IS AN HETEROMULTIMERIC ENZYME COMPOSED OF A
 CC PERIPHERAL CATALYTIC V1 COMPLEX (MAIN COMPONENTS: SUBUNITS A, B,
 CC C, D, E, AND F) ATTACHED TO AN INTEGRAL MEMBRANE V0 PROTON PORE
 CC COMPLEX (MAIN COMPONENT: THE PROTEOLIPID PROTEIN).
 CC -!- SUBCELLULAR LOCATION: PLASMA MEMBRANE OF THE RENAL INTERCALATED
 CC CELLS.
 CC -!- TISSUE SPECIFICITY: KIDNEY CORTEX AND MEDULLA.
 CC -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M88691; AAA30394.1; -
 DR PIR; C44138; C44138.
 DR InterPro; IPR003255; ATP_synt_B_C.
 DR InterPro; IPR000194; ATPase_alpha_beta.
 DR Pfam; PF00006; ATP_synt_ab; 1.
 DR Pfam; PF01038; ATP_synt_B_C; 1.
 DR PROSITE; PS00152; ATPase_ALPHA_BETA; 1.
 KW Hydrolase; ATP synthesis; Hydrogen ion transport; Multigene family.
 SQ SEQUENCE 513 AA; 56746 MW; 4C6DB84E12570A25 CRC64;

Query Match 5.2%; Score 102; DB 1; Length 513;
 Best Local Similarity 19.7%; Pred. No. 1.4;
 Matches 59; Conservative 49; Mismatches 108; Indels 84; Gaps 12;

QY 5 PREMFDCLKPIWRCRGLGYACLTILSRMKERVF-----CSRTCRITTTIQRGL 54
 DB 160 PEMIETGISPID-----VNNIARQKIPISAGLPHNEIAAQICR-----QAGL 206
 QY 55 ESKVQLGTQNVLDLKLVEWNNHFGIHFMRVSSDLFPFASHAKYGYTLEFAQSHLEEVGK 114
 DB 207 --VKK--SKAVLDY-----HDDNFAIVFAAMGVNM-----ETARFFKSDFEQNGT 247
 QY 115 LANK-----YNRLTMHFGQYTOIASPREVV-----DSAIRDL-AY 150
 DB 248 MGNVCLFLMLANDPTIERIITPRALTAETAEFLAYQCEKHVILVILTDMSSYAEALREVSAA 307
 QY 151 HDEILSRMKLNQKDAVLIILHGGTPEGKKTLDLDRFKNYQRLSDSVKARLVLENDV 210
 DB 308 REEVPCRRGFGPYMTDLATIVERAGRVEGRNGSITQI-----PILTMPNDI 355
 QY 211 SWSVODLLPLCOELNIPVLVDWHHNIIVPGTLREGSLDMLPIPTIRTWKGTQKOH 270
 DB 356 THPIPLDTGITEGQIYVDROLHNRQIYP-----PINVLPSLSRLMKSATCEGTRKDH 409

RESULT 13

VATB_CHICK STANDARD; PRT; 453 AA.
 ID VATB_CHICK
 AC P49712;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE VACUOLAR ATP SYNTHASE SUBUNIT B (EC 3.6.1.34) (V-ATPASE B SUBUNIT)
 DE (VACUOLAR PROTON PUMP B SUBUNIT).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-WHITE LECHORN: TISSUE-Bone;
RX MEDLINE=95369682; PubMed=7642089;
RA Bartkiewicz W., Hernando N., Reddy S.V., Roodman G.D., Baron R.;
RT "Characterization of the osteoclast vacuolar H(+)-ATPase B-subunit.";
RL Gene 160:157-164(1995).
CC -!- FUNCTION: NONCATALYTIC SUBUNIT OF THE PERIPHERAL V1 COMPLEX OF
CC VACUOLAR ATPASE. V-ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY
CC OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS.
CC -!- SUBUNIT: V-ATPASE IS AN HETEROMULTIMERIC ENZYME COMPOSED OF A
CC PERIPHERAL CATALYTIC V1 COMPLEX (MAIN COMPONENTS: SUBUNITS A, B,
CC C, D, E, AND F) ATTACHED TO AN INTEGRAL MEMBRANE V0 PROTON PORE
CC COMPLEX (MAIN COMPONENT: THE PROTEOLIPID PROTEIN).
CC -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U20766; AAA82983.1; -
DR InterPro: IPR003255; ATP_synt_B_c.
DR InterPro: IPR000194; ATPase_alpha_beta.
DR Pfam: PF00006; ATP_synt_ab; 1.
DR Pfam: PF01038; ATP_synt_B_c; 1.
DR Pfam: PF01038; ATP_synt_B_c; 1.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; 1.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; 1.
KW Hydrolyase; ATP synthetase; Hydrogen ion transport; Multigene family.
SQ SEQUENCE 453 AA; 50225 MW; 463885F93DFEC962 CRC64;

Query Match 5.1%; Score 101.5; DB 1; Length 453;
Best Local Similarity 19.2%; Pred. No. 1.3;
Matches 61; Conservative 55; Mismatches 111; Indels 91; Gaps 14;
QY 2 DHAP-----REMFCDLDRPIWRGR-----GYAC---LNTILSRMKERVF----- 39
DB 81 DRGPAVLAEDFLDIMGDPNQCRIYPEEMTQGISAGDGNSTARGOKPIFSAAGLPH 140
QY 40 ---CSRTCRITTIQDGLSVKOLGTQNVLDLIKLVENHNFHGMVSDLPFFASHA 96
DB 141 NEIAAQICR---QAGL--VKK--SKDMDYSE---ENFAVFAAMGVNM----- 179
QY 97 KYGYTFLEFAQSHLEVGKLANK-----YHRLTHMHPQYQTIASPREVVV 141
DB 180 ---ETAREFKSDFENGSMQNVCLFLNLANDPTIERTIPLALTTAEFLAYQCEKHVLY 236
QY 142 -----DSAIRDL-AYHDEILSRMKLNQKNDKDAVLIHLGCTFECKKTLDRFRKNY 192
DB 237 ILTDMSSVAEALREVSAREEVGRRGPGGYMTDLATYERACRVEGRNGSITQI---- 292
QY 193 QRLSDSVKARVLENDVSWSDVLLPLCQELNPLVDWHHHNIVPTGLREGSLDMLPL 252
DB 293 -----PILTMPNDITHPIPLDTGYITGQIVYVDRQLNRQIYP-----PINVLP 338
QY 253 IPTIRETRKGITOKOH 270
DB 339 LSRLMKSAIGEGWTRKH 356

RESULT 14
VATB_PYAB
ID VATB_PYAB STANDARD; PRT; 465 AA.
AC Q90X08;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE V-TYPE ATP SYNTHASE BETA CHAIN (EC 3.6.1.34) (V-TYPE ATPASE SUBUNIT
DE B).
GN ATPB OR PAB1186.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.

OX NCBI_TaxID=29292;
RN {1}
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAY;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
RT structure and evolution.";
RL Submitted (Jul-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC GRADIENT ACROSS THE MEMBRANE. THE ARCHAEL BETA CHAIN IS A
CC REGULATORY SUBUNIT.
CC -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; A3248288; CAB50665.1; -
DR InterPro: IPR003255; ATP_synt_B_c.
DR InterPro: IPR000194; ATPase_alpha_beta.
DR Pfam: PF00006; ATP_synt_ab; 1.
DR Pfam: PF01038; ATP_synt_B_c; 1.
DR Pfam: PS00152; ATPASE_ALPHA_BETA; 1.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; 1.
KW Hydrolyase; ATP synthetase; Hydrogen ion transport; Complete proteome.
SQ SEQUENCE 465 AA; 52224 MW; 873F500BD4C5EDEF CRC64;

Query Match 5.1%; Score 101.5; DB 1; Length 465;
Best Local Similarity 19.2%; Pred. No. 1.3;
Matches 59; Conservative 46; Mismatches 119; Indels 83; Gaps 12;
QY 27 LNTILSRMKERVFCSRTCRITTIQDGLSVKOLGTQNVLDLIKLVENHNFHGMVRS 86
DB 141 MNTLVRCQKLPIESG-----SGLPH-NQLAAQ-IARQAKVLGESEFAVVF---- 184
QY 87 SDLPFPASHAKYGYTLE---FAQSHLEEVGKLANK-----YHRLTHMHPG 128
DB 185 -----AAGITTYEAFNFFKSKFEETGAIERAVLFLNLADDDPAIERITPRMALTV 235
QY 129 QYQTIASPREVVV-----DSAIRDL-AYHDEILSRMKLNQKNDKDAVLIHLGCTFE 179
DB 236 EYLAFYDMQVLLVLDHMTNYCEALREISAREEVGRRGPGGYMTDLATYERAGRIR 295
QY 180 GKKETLDRFRKNYORLSVSKARVLENDVSWSDVLLPLCQELNPLVDWHHHNIVP 239
DB 296 GKKSITQM-----PILTMPDDITHPIPLDTGYITGQIVLSRELHRRGIYP 343
QY 240 GTLREGSLDMLPLIPTIRETRKGITOKOH-----YSES---ADPTAISGMKRR 286
DB 344 -----PIDVLPSSLRMKDGIGKGRTRDPOLAOOLYAAAYAEGRSLRDLVAVVGEAL 397
QY 287 AHSDRVF 293
DB 398 SETDKKY 404

RESULT 15
VATB_CAEEL
ID VATB_CAEEL STANDARD; PRT; 491 AA.
AC Q19626;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE VACUOLAR ATP SYNTHASE SUBUNIT B (EC 3.6.1.34) (V-ATPASE B
DE SUBUNIT) (VACUOLAR PROTON PUMP B SUBUNIT).
GN F20B6.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.


```
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: NAKADA=6
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 773 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-966-389-4

Query Match          4.5%; Score 88.5; DB 2; Length 773;
Best Local Similarity 20.2%; Pred. No. 1.5;
Matches 50; Conservative 41; Mismatches 85; Indels 71; Gaps 10;

QY 98 YGYTLEFAQSHL-EVVGKL-----ANKYNHRLTMHPGOYTQIASPREVVVDSAIRDLAYHD 152
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 461 YGCEILFETAREWEDLGAVILPKGNKFCINCVTGPDEYALVDNNAYTNMAKNMLEYAY 520

QY 153 EILSRMKLNEQLNKDAVLIILHGGTFEGKKEITLDRFRKNYORLSDSVKARLVLEDD-VS 211
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 521 DIANKM-----KKEVPQKYOK-----VASKLNKDEEIVA 550

QY 212 W---SVQDLLPLCOELNI-----PLVLDWHHHNIVPGTLREG 245
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 551 WKKAADNMVLPYSKELDILIPQDDSFYKERTVDEIPEDQFPLLLHHVLYNIYRYQICK- 609

QY 246 SLDLMLPIPTIRETWRTKGITOKHYSE-----SADPTAISGMKRR-AHSDRVFDEPP 297
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 610 QPDVLLMFLOREKFTKDELKKNYDYEPITTHDSSLSPAIFSIILANEIGYTDKAYKYM 669

QY 298 CDPTMDL 304
   |||
Db 670 MTARMDL 676

RESULT 7
US-09-103-509-4
; Sequence 4, Application US/09103509
; Patent No. 5876975
; GENERAL INFORMATION:
; APPLICANT: Tetsuya NAKADA
; APPLICANT: Michio KUBOTA
; APPLICANT: Hiroto CHAEN
; APPLICANT: Toshio MIYAKE
; TITLE OF INVENTION: TREHALOSE PHOSPHORYLASE, ITS PREPARATION
; TITLE OF INVENTION: AND USES
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,509
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/966,389
; FILING DATE: 07-NOV-1997
; APPLICATION NUMBER: JP 311,232/1996
```

```
; FILING DATE: 8-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 61,716/1997
; FILING DATE: 3-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: NAKADA=6
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 773 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-103-509-4

Query Match          4.5%; Score 88.5; DB 2; Length 773;
Best Local Similarity 20.2%; Pred. No. 1.5;
Matches 50; Conservative 41; Mismatches 85; Indels 71; Gaps 10;

QY 98 YGYTLEFAQSHL-EVVGKL-----ANKYNHRLTMHPGOYTQIASPREVVVDSAIRDLAYHD 152
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 461 YGCEILFETAREWEDLGAVILPKGNKFCINCVTGPDEYALVDNNAYTNMAKNMLEYAY 520

QY 153 EILSRMKLNEQLNKDAVLIILHGGTFEGKKEITLDRFRKNYORLSDSVKARLVLEDD-VS 211
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 521 DIANKM-----KKEVPQKYOK-----VASKLNKDEEIVA 550

QY 212 W---SVQDLLPLCOELNI-----PLVLDWHHHNIVPGTLREG 245
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 551 WKKAADNMVLPYSKELDILIPQDDSFYKERTVDEIPEDQFPLLLHHVLYNIYRYQICK- 609

QY 246 SLDLMLPIPTIRETWRTKGITOKHYSE-----SADPTAISGMKRR-AHSDRVFDEPP 297
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 610 QPDVLLMFLOREKFTKDELKKNYDYEPITTHDSSLSPAIFSIILANEIGYTDKAYKYM 669

QY 298 CDPTMDL 304
   |||
Db 670 MTARMDL 676

RESULT 8
US-09-102-644-4
; Sequence 4, Application US/09102644
; Patent No. 5910436
; GENERAL INFORMATION:
; APPLICANT: Tetsuya NAKADA
; APPLICANT: Michio KUBOTA
; APPLICANT: Hiroto CHAEN
; APPLICANT: Toshio MIYAKE
; TITLE OF INVENTION: TREHALOSE PHOSPHORYLASE, ITS PREPARATION
; TITLE OF INVENTION: AND USES
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/102,644
; FILING DATE:
; CLASSIFICATION:
```


;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/319,866
;; FILING DATE: 7-OCT-1994
;; CLASSIFICATION: 514
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER:
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Granahan, Patricia
;; REGISTRATION NUMBER: 32,227
;; REFERENCE/DOCKET NUMBER: CSHL94-03
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617) 861-6240
;; TELEFAX: (617) 861-9540
;; INFORMATION FOR SEQ ID NO: 12:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1144 amino acids
;; TYPE: amino acid
;; STRANDEDNESS:
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-319-866-12

Query Match 4.4%; Score 88; DB 2; Length 1144;
Best Local Similarity 17.3%; Pred. No. 3.3;
Matches 46; Conservative 38; Mismatches 90; Indels 92; Gaps 9;

QY 37 RVFCSTCRITTIQDGL-----ESVKQLGTQNVLDLIKLVENHNFG 79
Db 562 KVVCMDOYKASTLEEOLLLVVTFTFGNGDCPSNGQTLKK-----SLFMLRELNHTFR 614
QY 80 IHFMVSSDLFPFASHAKYGYTLEFAQSHL-----EEVG----- 113
Db 615 YAVFGLGSSMYP--QCAFAHDIQKLSHGASQLAPTGEDELSCQEDAFRSWAVQTFR 672
QY 114 -----KLANKYNHRLTMHPGQYTOIASPREVVVDSAIRDLAYHDEILSRMK 159
Db 673 AACETFDVRSKHHIQIPKRFTSNATWEPQOYRLIQSPPEPLDNLRALSSIHAKNVFTMLRK 732
QY 160 LNEQLNKD-----AVLIHLGGTFEGKK-----ETLDRFRKNYORLSDSVKARL----- 203
Db 733 SQONLOSEKSSRTLLVQL--TFEGSRGSPYLPGEHLGIFPGNOTALVOGILERVVDCTP 790
QY 204 -----VLENDVSWSQDILLPLC 221
Db 791 PHQTVCLEVLDSEGSYVWKDKRLPPC 816

RESULT 13
US-09-123-708-2
; Sequence 2, Application US/09123708
; Patent No. 6146867
; GENERAL INFORMATION:
; APPLICANT: SCHRADER, Juergen
; APPLICANT: GODECKE, Axel
; TITLE OF INVENTION: DNA EXPRESSION VECTORS FOR USE IN GENE THERAPEUTIC
; TITLE OF INVENTION: TREATMENT OF VASCULAR DISORDERS
; FILE REFERENCE: 511169-2003
; CURRENT APPLICATION NUMBER: US/09/123,708
; CURRENT FILING DATE: 1998-07-28
; EARLIER APPLICATION NUMBER: 08/553,503
; EARLIER FILING DATE: 1996-03-01
; EARLIER APPLICATION NUMBER: P411402.8
; EARLIER FILING DATE: 1994-03-31
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1144
; TYPE: PRT
; ORGANISM: Cytomegalovirus
US-09-123-708-2

Query Match 4.4%; Score 88; DB 4; Length 1144;
Best Local Similarity 17.3%; Pred. No. 3.3;
Matches 46; Conservative 38; Mismatches 90; Indels 92; Gaps 9;

QY 37 RVFCSTCRITTIQDGL-----ESVKQLGTQNVLDLIKLVENHNFG 79
Db 562 KVVCMDOYKASTLEEOLLLVVTFTFGNGDCPSNGQTLKK-----SLFMLRELNHTFR 614
QY 80 IHFMVSSDLFPFASHAKYGYTLEFAQSHL-----EEVG----- 113
Db 615 YAVFGLGSSMYP--QCAFAHDIQKLSHGASQLAPTGEDELSCQEDAFRSWAVQTFR 672
QY 114 -----KLANKYNHRLTMHPGQYTOIASPREVVVDSAIRDLAYHDEILSRMK 159
Db 673 AACETFDVRSKHHIQIPKRFTSNATWEPQOYRLIQSPPEPLDNLRALSSIHAKNVFTMLRK 732
QY 160 LNEQLNKD-----AVLIHLGGTFEGKK-----ETLDRFRKNYORLSDSVKARL----- 203
Db 733 SQONLOSEKSSRTLLVQL--TFEGSRGSPYLPGEHLGIFPGNOTALVOGILERVVDCTP 790
QY 204 -----VLENDVSWSQDILLPLC 221
Db 791 PHQTVCLEVLDSEGSYVWKDKRLPPC 816

RESULT 14
US-09-123-624-2
; Sequence 2, Application US/09123624
; Patent No. 6149936
; GENERAL INFORMATION:
; APPLICANT: SCHRADER, Jorgen
; APPLICANT: GODECKE, Axel
; TITLE OF INVENTION: DNA EXPRESSION VECTORS FOR USE IN THE GENE THERAPEUTIC
; TITLE OF INVENTION: TREATMENT OF VASCULAR DISORDERS
; FILE REFERENCE: 511169-2004
; CURRENT APPLICATION NUMBER: US/09/123,624
; CURRENT FILING DATE: 1998-07-28
; PRIOR APPLICATION NUMBER: 08/553,503
; PRIOR FILING DATE: 1996-03-01
; PRIOR APPLICATION NUMBER: 4411402.8
; PRIOR FILING DATE: 1994-03-31
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1144
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-123-624-2

Query Match 4.4%; Score 88; DB 4; Length 1144;
Best Local Similarity 17.3%; Pred. No. 3.3;
Matches 46; Conservative 38; Mismatches 90; Indels 92; Gaps 9;

QY 37 RVFCSTCRITTIQDGL-----ESVKQLGTQNVLDLIKLVENHNFG 79
Db 562 KVVCMDOYKASTLEEOLLLVVTFTFGNGDCPSNGQTLKK-----SLFMLRELNHTFR 614
QY 80 IHFMVSSDLFPFASHAKYGYTLEFAQSHL-----EEVG----- 113
Db 615 YAVFGLGSSMYP--QCAFAHDIQKLSHGASQLAPTGEDELSCQEDAFRSWAVQTFR 672
QY 114 -----KLANKYNHRLTMHPGQYTOIASPREVVVDSAIRDLAYHDEILSRMK 159
Db 673 AACETFDVRSKHHIQIPKRFTSNATWEPQOYRLIQSPPEPLDNLRALSSIHAKNVFTMLRK 732
QY 160 LNEQLNKD-----AVLIHLGGTFEGKK-----ETLDRFRKNYORLSDSVKARL----- 203
Db 733 SQONLOSEKSSRTLLVQL--TFEGSRGSPYLPGEHLGIFPGNOTALVOGILERVVDCTP 790
QY 204 -----VLENDVSWSQDILLPLC 221
Db 791 PHQTVCLEVLDSEGSYVWKDKRLPPC 816

Db 791 PHQTVCLVLEDSGSYVWKRLPPC 816

RESULT 15

US-08-560-005-2
; Sequence 2, Application US/08560005
; Patent No. 6001354
; GENERAL INFORMATION:
; APPLICANT: Pot, David A.
; APPLICANT: Williams, Lewis T.
; APPLICANT: Jefferson, Anne Bennett
; APPLICANT: Majerus, Philip W.
; TITLE OF INVENTION: No. 6001354el Grb2 Associating Protein and Nucleic
; TITLE OF INVENTION: Acids Encoding Therefor
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/560,005
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-0624000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 976 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-560-005-2

Query Match 4.4%; Score 87; DB 3; Length 976;

Best Local Similarity 20.4%; Pred. No. 3.3;
Matches 79; Conservative 53; Mismatches 130; Indels 126; Gaps 18;

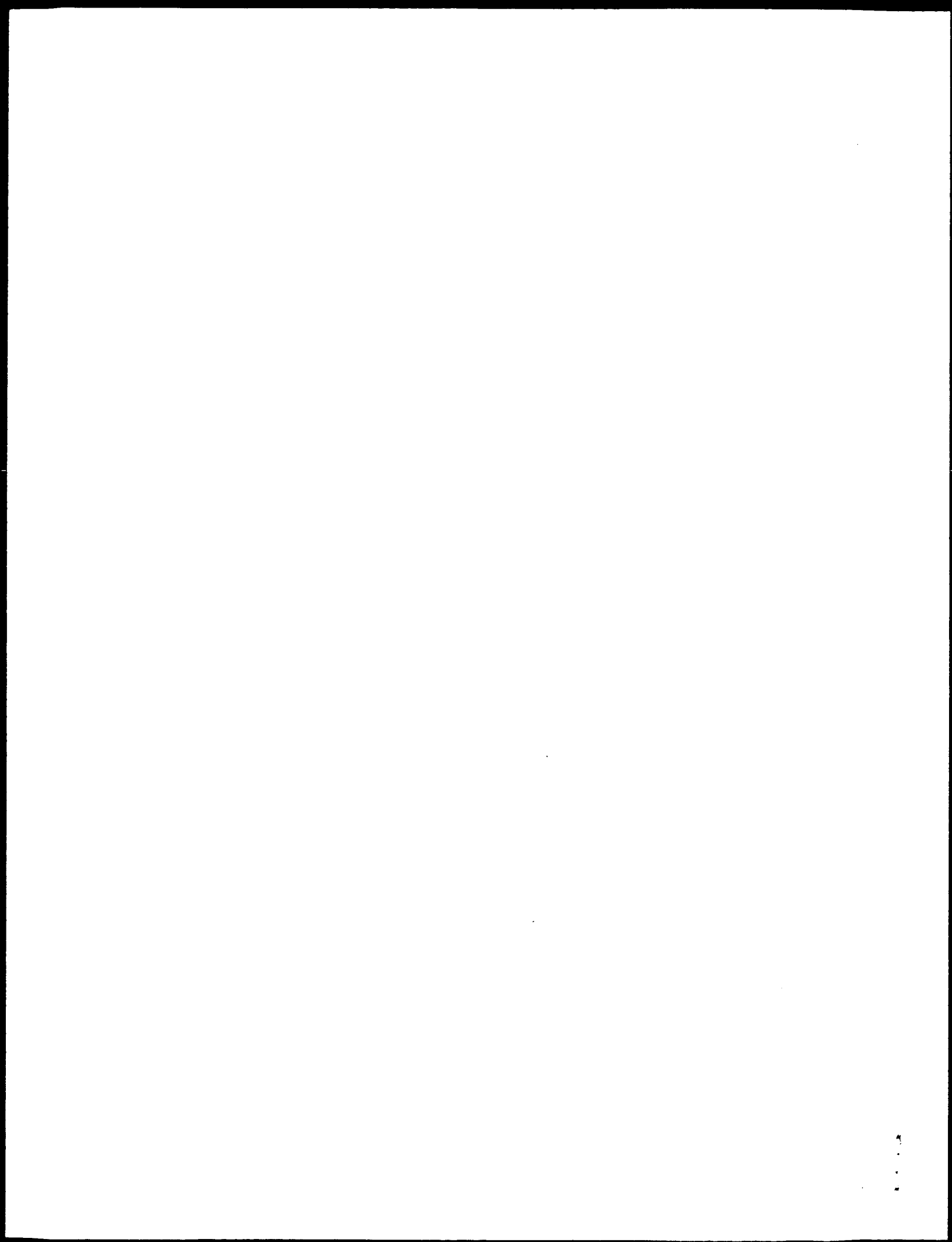
QY 24 YACLNTILRSKMERVFCRTRCITT--IQRDGLSVKQLGTQNVLDLIKLVENHNFGIH 81
Db 477 YGSTDIMTSDHSPVFTEACVTSQFVSKNGPGTVDSQGG-----IE 519
QY 82 FMRVSSDLPPFASHAKYGYTLEFAOSHLEEV-----GKLANKYNHRLTMHPG 128
Db 520 FLRCYATL---RTKSQTKFYLEFHSCLSEFVKSOGEENEGEGELVYKFGETLP----- 572
QY 129 OYQTJASPREVVVD-----SAIRDLAYHDEILS-RMKLNE-OLNKDAVLIH--LGG 176
Db 573 KIKPIISDEYLLDQHILLISIKSSDSSEYGGCGALREATETOLPIYTPLTTHGELTG 632
QY 177 TFEGRKETLDRFRKNYQRLSDSVKARLVLENDVVS-----WSVODLLPL 220
Db 633 HFQGEIKLQTSQGTREKLYDFVKT---ERDESSGPKTLKSLTSHDPMQWEVTSRAPP 688
QY 221 COELNIPVLVDHHHHNIVPGTLREGSLDMLPIPIRETWTWKGITQKHYSADPTAI 280
Db 689 CSGSSITEIFINMYGVGP-----FGPPMP-----LHVKOTLSPDQOQTAW 729
QY 281 SGMKRAHSDRVDFPPCPDPTMDLMEAKEKEQAVFELCRRYELQNPCCP--LEIMGP 338

Db 730 S-----YDQPPKDSPLG-----PCRG-----ESPPTPGQPPLSPK- 760

QY 339 DQTRDGYYPGAEKRLTARKRRSRKEEV 366

Db 761 -----KFLPSTANRGLPPRTQESRPSDL 783

Search completed: January 15, 2002, 13:51:04
Job time: 333 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 14:07:26 ; Search time 45.07 Seconds
(without alignments)
239.171 Million cell updates/sec

Title: US-09-724-296-39

Perfect score: 1527

Sequence: 1 QLGLVCLTVGPEVFRFTVL.....VEAKGKEAIAALRLMAPFK 294

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	303	19.8	320	1 YWJD_BACSU	P45864 bacillus su
2	90.5	5.9	457	1 LEFA_NPVOP	O10340 orgyia pseu
3	90	5.9	488	1 DHAL_PSESP	P33008 pseudomonas
4	90	5.9	607	1 NODM_RHLV	P08633 rhizobium l
5	89	5.8	139	1 PUAC_STRLP	P13249 streptomyc
6	89	5.8	2472	1 NCR2_MOUSE	Q9WU42 mus musculu
7	88	5.8	537	1 P4H2_MOUSE	O60716 mus musculu
8	86	5.6	1110	1 PKNK_MYCTU	P95078 mycobacteri
9	85.5	5.6	653	1 APPI_MOUSE	Q03157 mus musculu
10	84.5	5.5	304	1 HEY1_CANFA	Q9TSZ2 canis fami
11	84.5	5.5	616	1 RFX5_HUMAN	P48382 homo sapien
12	84	5.5	378	1 RPSD_CLOAB	P33656 clostridium
13	82.5	5.4	271	1 TRPA_MYCIT	O68906 mycobacteri
14	82.5	5.4	650	1 APPI_HUMAN	P51693 homo sapien
15	82.5	5.4	1160	1 DP3A_ECOLI	P10443 escherichia
16	82.5	5.4	3567	1 ERY2_SACER	Q03132 saccharopol
17	82	5.4	435	1 COBB_RHOCA	O68108 rhodobacter
18	81.5	5.3	967	1 HEPA_ECOLI	P23852 escherichia
19	81	5.3	506	1 GAG_SIVM1	P05894 simian immu
20	81	5.3	607	1 DNAK_STRPN	P95829 streptococ
21	81	5.3	974	1 ATXB_LEIDO	P12522 leishmania
22	81	5.3	1172	1 SYK2_MYCTU	P94974 mycobacteri
23	80.5	5.3	476	1 PCCB_RHOER	O06101 rhodococcus
24	80.5	5.3	859	1 OBP_HSVBC	P52377 bovine herp
25	80	5.2	307	1 BLAC_MYCTU	Q10670 mycobacteri
26	80	5.2	404	1 TRPB_BACST	P19868 bacillus st
27	80	5.2	612	1 UN37_CAEEL	Q02482 caenorhabdi
28	80	5.2	1002	1 TNPA_ECOLI	Q00937 escherichia
29	80	5.2	1257	1 KPBA_CAEEL	P34335 caenorhabdi
30	79.5	5.2	159	1 ISPF_MYCTU	P96863 mycobacteri
31	79.5	5.2	406	1 TRPB_CAUCR	P12290 caulobacter
32	79.5	5.2	474	1 VNFK_AZOCH	P15334 azotobacter
33	79.5	5.2	492	1 NYLA_FLASK	P13397 flavobacter

RESULT

```

YWJD_BACSU
ID YWJD_BACSU STANDARD; PRT: 320 AA.
AC P45864:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 36.9 KDA PROTEIN IN ACDA 5'REGION.
GN YWJD.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Glaser P., de la Fuente V., Danchin A.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 249782; CA889865.1; -
DR EMBL; 259123; CAB15748.1; -
DR Subtilist; BG11309; ywjd.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 320 AA; 36896 MW; 2FD35A710648E002 CRC64;

```

Query Match 19.8%; Score 303; DB 1; Length 320;

Best Local Similarity 30.2%; Pred. No. 4.9e-18;

Matches 90; Conservative 50; Mismatches 116; Indels 42; Gaps 11;

```

QY 16 RIVTLSRYRALSPAREAKLLDLYSSNIKTGAADYCAADIRLYRSLSPML---D 72
DB 21 KTLTFARYSKLSKTERKEALLTVTKANLRNTMRTLHYITGHGIPLYRFSSTVPLATHPD 80
QY 73 LAGDDTGAAVLTHLAPOLLEAGHAFTDAGVRLMLHPEQFVLNSDRPEVRESSVRMSAH 132
DB 81 VWMD-----FVTFQKEFEIGELVKYTHQLTSTSPKESVTKNAVTDMAH 135
QY 133 ARVMDGLGLA-RTPWNLL--LHGKGGRGAEALALIPDLDPVRLRLGLENDERAYSPA 189
DB 136 YRMLEAMGIADRSVINIHIGGAYGNKDTATAQHONIKOLPOEIKERMITLENDDKTYTTE 195
QY 190 ELLPICEATGTPLVFDAAHHVVDKLPDQEDSVREWVLRARATW-----QPPEQVYHL 244
DB 196 ETLVQVCEQEDVPFVDFHHFYANP--DDHADLNV--LPRMIKTWERIGLOPK----VHL 247
QY 245 SNGIEGPDRRHSHLHIADEFPASVADVPOTE-----VEAKGKEAIAALRLM 290

```

ALIGNMENTS

```

34 79 5.2 470 1 SELA_MOOTH 033277 moorella th
35 79 5.2 659 1 CAOP_CAEEL P34355 caenorhabdi
36 79 5.2 1075 1 NRC3_MOUSE P97305 mus musculu
37 79 5.2 1184 1 FBL2_HUMAN P98095 homo sapien
38 79 5.2 2346 1 COAL_HUMAN Q13085 homo sapien
39 78.5 5.1 271 1 DAPB_MYCBO P46829 mycobacteri
40 78.5 5.1 393 1 XYLA_ACTMI P12851 actinoplan
41 78.5 5.1 1136 1 TIE1_BOVIN Q06805 bos taurus
42 78.5 5.1 3491 1 ERY1_SACER Q03131 saccharopol
43 78 5.1 956 1 DPOL_DEIRA P32027 deinococcus
44 74 5.1 3519 1 OLS6_STRAT Q07017 streptomyce
45 77.5 5.1 304 1 HEY1_HUMAN Q95753 homo sapien

```


Db 422 LAESLGHVLEHGDAGRYVLTSLIDEIHE-----NDDRIAVV 459
 QY 122 -----RESSVRAMSAHARVMDG-----LGLARTPNMLLLHGGKGGGAEIAALIPDLPDP 172
 Db 460 IDWHRVSDSRQAALGFLONGCHHLLQIVTSWSRAGLPVGRVIRIGDELAEI-----DS 514
 QY 173 VRLRLGLENDERA-----YSPAELLPICEAT----- 198
 Db 515 AALR--FDTDEAALLNDAGGLRLPRADVQALTTSTDGMAAALRLAALSRLGGGDATQLL 572
 QY 199 -CTPLVFDAAHHVVDKLPDQEDPSVREWLRA-----RATWQPP 237
 Db 573 RGLSGASDVIIHEFLSENLDLTLEPELREFLIVASVTERTCGLASALAGIINGRAMLEEA 632
 QY 238 EMOVYHLSNGIEGPDQRHSHLIADF 263
 Db 633 EHRGLFLORTEDDPNFRFQMFADF 658

 RESULT 9
 ID APPL_MOUSE STANDARD; PRT; 653 AA.
 AC Q03157;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE AMYLOID-LIKE PROTEIN 1 PRECURSOR (APLP).
 GN APLP1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=93066322; PubMed=1279693;
 RA Wasco W., Bupp K., Magendanz M., Gusek J.F., Tanzi R.E.,
 RA Solomon F.;
 RT "Identification of a mouse brain cDNA that encodes a protein related
 RT to the Alzheimer disease-associated amyloid beta protein precursor.";
 RT Proc. Natl. Acad. Sci. U.S.A. 89:10758-10762(1992).
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. MAY BE PROCESSED
 CC IN THE GOLGI COMPLEX; HOWEVER, IT IS NOT YET CLEAR WHETHER APLP
 CC IS SECRETED.
 CC -!- SIMILARITY: BELONGS TO THE APP FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L04538; AAA37247.1; -;
 DR PIR; A46362; A46362.
 DR MGD; MGI:88046; Aapl1.
 DR InterPro; IPR001868; A4_APP.
 DR Pfam; PF02177; A4_EXTRA; 1.
 DR PRINTS; PR0203; AMYLOIDA4.
 DR SMART; SM00006; A4_EXTRA; 1.
 DR PROSITE; PS00319; A4_EXTRA; 1.
 DR PROSITE; PS00320; A4_INTRA; 1.
 DR Glycoprotein; Transmembrane; Signal.
 KW SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 653
 FT DOMAIN 21 583
 FT TRANSMEM 584 606
 FT DOMAIN 607 653
 FT DOMAIN 643 646
 FT DOMAIN 263 271
 FT CARBOHYD 464 464
 FT CARBOHYD 534 554
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).

 SQ SEQUENCE 653 AA; 72751 MW; 56516DC3EA40E4B0 CRC64;

 Query Match 5.6%; Score 85.5; DB 1; Length 653;
 Best Local Similarity 21.2%; Pred. NO.13;
 Matches 55; Conservative 22; Mismatches 83; Indels 99; Gaps 10;

 QY 53 CAADIRLRYLRLSSSLPMLDLA-----GDDTGAAVLTH--LAPQLLEAG 94
 Db 83 CLLDLPQVLEVCYCHQMPGLHARVEQAQAIPMERWCCTGTRSGCAHPHVEVPPHCLPG 142
 QY 95 HATDAGVRLMLLMIPEQFIVLNSDRPEVRESSVR-----AMSAHARVMDGLGLARTPN 147
 Db 143 EYFSEA-----LLVPEGRFLHQRMDQCESSTRKHQEAQACSSQGLLIHSG----- 191
 QY 148 LLLHGGKGGKGAELAAALPDIPDPVRLRLGLENDERAYSPAELLPICEATGTPLVFD 207
 Db 192 MLLPGSDRFRGVYVCCPP-----PATPNPSGMAAG----- 223
 QY 208 HHVVDKLPDQEDPSVREWLRAATWQPEVQVHLSNGIEGPDQRHSHLIADPPSA- 266
 Db 224 -----DPSTRSWPLGGA-----EGGEDEEE---VESFPQPV 252
 QY 267 ---YADVPEQIEVEAKGKEE 282
 Db 253 DDYFVEPPQAEHEEEEEEE 271

 RESULT 10
 ID HEX1_CANFA STANDARD; PRT; 304 AA.
 AC Q0TS22;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HAIRY/ENHANCER-OF-SPLIT RELATED WITH YRP MOTIF 1 (HAIRY AND ENHANCER
 DE OF SPLIT RELATED-1) (HESR-1).
 GN HEY1 OR HESR1.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Thyroid;
 RA Pichon B., Mercan D., Pouillon V., Christophe-Hobertus C.,
 RA Christophe D.;
 RT "Functional cloning of nuclear proteins and nuclear targeting
 RT sequences.";
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
 CC BHLH PROTEIN (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
 CC TRANSCRIPTION FACTORS. "HAIRY" SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AJ388551; CAB65543.1; -;
 DR InterPro; IPR003015; HLH_Myc.
 DR InterPro; IPR001092; HLH_dim.
 DR Pfam; PF00010; HLH; 1.
 DR SMART; SM00353; HLH; 1.
 DR SMART; SM00511; ORANGE; 1.
 DR PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
 KW Transcription regulation; DNA-binding; Nuclear protein.


```
FT DNA_BIND 50 62 BASIC DOMAIN.
FT DOMAIN 63 105 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 294 297 YRPW MOTIF (REQUIRED FOR ACTIVITY)
FT (POTENTIAL).
SQ SEQUENCE 304 AA: 32471 MW: 191D5BD42970F412 CRC64:

Query Match 5.5%; Score 84.5; DB 1; Length 304;
Best Local Similarity 24.7%; Pred. No. 6.1;
Matches 58; Conservative 23; Mismatches 79; Indels 75; Gaps 13:

QY 19 TISRYALPAERRE-----AKLLDLYSSNIXTLR--GAADYCAAHDIRL-YRLSSS 66
DB 68 SLSELRVPSAFKOGSKAKLEAILQTVHKLMLTAGGKGYFDAHALAMDYR---- 123
QY 67 LFPMLDAGDGTGAAYVTLHAPOLLEAGHAFTDAGVRLMLHPPEQFIVLNSDRPREVRESSV 126
DB 124 -----SLGFRCLAEVARYLS--IIEGLDASDPLRVRLVSH-----LNRYASQ-REA-- 167
QY 127 RAMSAHARVMDGLGLARTPNW-----LLLLHGKGGRG-----AELA 163
DB 168 -ASGAHA-----GLGHLPGWSAFGHHPHVAHPLLLPQSGHGTGTSAPTDPHQRGLA 220
QY 164 ALIPDLP-----DPVRLRLGLENDERAYSPAELLPICEATGTPLVFDAHH 208
DB 221 AAHPEAPALRAPSGGLGPV---LPVVTASKSLPPLLSSVASLSAFAFPFGSGFH 272

RESULT 11
REFX5_HUMAN STANDARD; PRT; 616 AA.
AC P48382;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA-BINDING PROTEIN RFX5 (REGULATORY FACTOR X SUBUNIT 5).
GN RFX5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=95262896; PubMed=7744245;
RA Steinle V., Durand B., Barras E., Zufferey M., Hadam M.R.,
RA Mach B., Reith W.;
RT "A novel DNA-binding regulatory factor is mutated in primary MHC
RT class II deficiency (bare lymphocyte syndrome).";
RL Genes Dev. 9:1021-1032(1995).
RN [2]
RP PARTIAL SEQUENCE.
RC TISSUE=Lymphoblast;
RX MEDLINE=99170284; PubMed=10072068;
RA Nagarajan U.M., Louis-Plence P., Desandro A., Nilsen R., Bushey A.,
RA Boss J.M.;
RT "RFX-B is the gene responsible for the most common cause of the bare
RT lymphocyte syndrome, an MHC class II immunodeficiency.";
RL Immunity 10:153-162(1999).
RN [3]
RP ERRATUM.
RA Nagarajan U.M., Louis-Plence P., Desandro A., Nilsen R., Bushey A.,
RA Boss J.M.;
RL Immunity 10:399-399(1999).
RN [4]
RN CHARACTERIZATION.
RX MEDLINE=2042030; PubMed=10779326;
RA Villard J., Peretti M., Masternak K., Barras E., Caretti G.,
RA Mantovani R., Reith W.;
RT "A functionally essential domain of RFX5 mediates activation of major
RT histocompatibility complex class II promoters by promoting
RT cooperative binding between RFX and NF-Y.";
RL Mol. Cell. Biol. 20:3364-3376(2000).
RN [5]
RT FUNCTION: ACTIVATES TRANSCRIPTION FROM CLASS II MHC PROMOTERS.
CC
```

```
CC RECOGNIZES X-BOXES. MEDIATES COOPERATIVE BINDING BETWEEN RFX AND
CC NF-Y. RFX BINDS THE X1 BOX OF MHC-II PROMOTERS.
CC -1- SUBUNIT RFX CONSISTS OF AT LEAST THREE DIFFERENT SUBUNITS: RFXAP,
CC RFX5 AND RFX-B/RFXANK; WITH EACH SUBUNIT REPRESENTING A SEPARATE
CC COMPLEMENTATION GROUP. RFX FORMS COOPERATIVE DNA BINDING COMPLEXES
CC WITH X2BP AND CBF/NF-Y. RFX ASSOCIATES WITH CIITA TO FORM AN
CC ACTIVE TRANSCRIPTIONAL COMPLEX.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: UBIQUITOUS.
CC -1- DOMAIN: THE N-TERMINUS IS REQUIRED FOR ITS ASSOCIATION WITH RFXANK
CC AND RFXAP, FOR ASSEMBLY OF THE RFX COMPLEX, AND FOR BINDING OF
CC THIS COMPLEX TO ITS X BOX TARGET SITE IN THE MHC-II PROMOTER. THE
CC C-TERMINUS MEDIATES COOPERATIVE BINDING BETWEEN THE RFX COMPLEX
CC AND NF-Y.
CC -1- PTM: PHOSPHORYLATED.
CC -1- DISEASE: DEFECTS IN RFX5 ARE A CAUSE OF HEREDITARY MHC CLASS II
CC DEFICIENCY (ALSO KNOWN AS BARE LYMPHOCYTE SYNDROME (BLS) OR HLA
CC CLASS II-DEFICIENT COMBINED IMMUNODEFICIENCY); A FORM OF SEVERE
CC COMBINED IMMUNODEFICIENCY DISEASE (SCID). RFX5 IS LINKED WITH
CC BLS COMPLEMENTATION GROUPS C.
CC -1- SIMILARITY: BELONGS TO THE RFX FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

RA Sauer U., Treuner A., Buchholz M., Santangelo J.D., Durre P.:
RT "Sporulation and primary sigma factor homologous genes in Clostridium
RL acetobutylicum."
RN J. Bacteriol. 176:6572-6582(1994).
RP
RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum."
RL J. Bacteriol. 183:4823-4838(2001).
CC -!- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES
CC ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND
CC THEN IS RELEASED. THIS IS THE PRIMARY SIGMA-FACTOR OF THIS
CC BACTERIA.
CC -!- SIMILARITY: BELONGS TO THE SIGMA-70 FACTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z23080; CAA80625.1; -
CC EMBL; AE007642; AAK79271.1; -
CC PIR; S34307; S34307.
CC HSSP; P00579; 1STG.
CC InterPro; IPR000943; Sigma_70.
CC Pfam; PF00140; sigma70.1.
CC PRINTS; PR00046; SIGMA70FCT.
CC PROSITE; PS00715; SIGMA70.1; 1.
CC PROSITE; PS00716; SIGMA70.2; 1.
CC Transcription regulation; Sigma factor; DNA-directed RNA polymerase;
CC DNA-binding.
CC DOMAIN 169 182 POLYMERASE CORE BINDING (POTENTIAL).
CC DNA_BIND 339 358 H-T-H MOTIF (BY SIMILARITY).
CC FT SEQUENCE 378 AA; 43260 MW; 13821F827E5D50B8 CRC64;
SQ
Query Match 5.5%; Score 84; DB 1; Length 378;
Best Local Similarity 22.5%; Pred. No. 8.8;
Matches 62; Conservative 45; Mismatches 95; Indels 74; Gaps 14;
QY 58 IRLYRLSSLPFMI-----DLA-----GD-----DTGAAVLTHLAPLLEAGHAFT 98
DB 109 VRMYLKETGKVPLLSPSEEIDLAQRKNGDRSARKKLAEANLRVSVIAKRYVGRGMLFL 168
QY 99 D-----AGVRLLMHPPOFTVLNSDRPE-----VRESSVRAMSAHARVMDGLGARTPWNL 148
DB 169 DLIOEGNGLLIKAVEKDFKGFSTYATWVWIKOATRAIOAQRTI-----RIPVHM 222
QY 149 -----LLHGGGKGGGAELALIPDL-DPVRRLRLGLENDERAYSPAEL-LP 193
DB 223 VETINKLIRVSRQLLQELGRPOPEEIAKIM-DMPVDKVR-----EIMKTAQEPVSLETP 276
QY 194 ICEATGTPVFDAAHHVHDKLPDQEDPS-----VREWVLRARATWQPPWQVHVL 244
DB 277 IGEE-----DSH-----LGFDFIPDEAPADAAAFMKLEQLLKINTLTTPREKVLRL 327
QY 245 SNGTEGPDRRHSHLIADFPFSAYADVPOIEVEAKGK 280
DB 328 RFLGDDGRARTLEEVGFEFNVNTRIRIQIEAKALRK 363
RESULT 13
TRPA_MYCIT
ID TRPA_MYCIT STANDARD; PRT; 271 AA.

```

```

AC 068906;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRYPTOPHAN SYNTHASE ALPHA CHAIN (EC 4.2.1.20).
GN TRPA.
OS Mycobacterium intracellulare.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1767;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BATTY;
RA Alavi M.R., Rouse D.A., Morris S.L.;
RT "Nucleotide sequence and functional analysis of the tryptophan
RT synthase genes of Mycobacterium intracellulare";
RL Submitted (APR-1998) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: THE ALPHA SUBUNIT IS RESPONSIBLE FOR THE ALDOL CLEAVAGE
CC OF INDOLEGLYCEROL PHOSPHATE TO INDOLE AND GLYCERALDEHYDE 3-
CC PHOSPHATE.
CC -!- CATALYTIC ACTIVITY: L-SERINE + L-(INDOL-3-YL)GLYCEROL 3-PHOSPHATE
CC = L-TRYPTOPHAN + GLYCERALDEHYDE 3-PHOSPHATE + H(2)O.
CC -!- PATHWAY: LAST (FIFTH) STEP IN BIOSYNTHESIS OF TRYPTOPHAN.
CC -!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE TRPA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF057042; AAC17135.1; -
CC InterPro; IPR003009; FMN enzyme.
CC InterPro; IPR002028; TRP_synthase_alpha.
CC Pfam; PF00290; trp_synth; 1.
CC ProDom; PD001335; TRP_synthase_alpha; 1.
CC PROSITE; PS00167; TRP_SYNTHASE_ALPHA; 1.
CC Tryptophan biosynthesis; Lyase.
CC KW SEQUENCE 271 AA; 27804 MW; BAE0C5BC0788318 CRC64;
SQ

```

```

Query Match 5.4%; Score 82.5; DB 1; Length 271;
Best Local Similarity 21.9%; Pred. No. 7.7;
Matches 32; Conservative 19; Mismatches 40; Indels 55; Gaps 4;
QY 64 SSSLPMLDLAGDDTGAAVLTHLAP-----QLLEAGHAFTDAGVR 103
DB 8 ASRLGPIFDVCRDGRALIGLYPTGYDVTTSVIANVALVESGCDIVEGVPSDPM- 66
QY 104 LLMHPEQFTVLNSDRPEVRESSVRAMSAHARVMD-----GLGLARTPWNL 151
DB 67 -----DGFITVKATEAALHGGVVRVDRDTLAAVEALISAAGHAVVNTVNPVLR 113
QY 152 HG-----GKGGGGAELALIPD 168
DB 114 YGIDAFARILASAGGYGLITPDLPD 139
RESULT 14
APPL_HUMAN
ID APPL_HUMAN STANDARD; PRT; 650 AA.
AC P51693; O00113;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE AMYLOID-LIKE PROTEIN 1 PRECURSOR (APLP).
GN APLP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98088960; PubMed=9428684;
RA Paliga K., Peraus G., Kreger S., Durrwang U., Hesse L., Multhaup G.,
RA Masters C.L., Beyreuther K., Weidemann A.;
RT "Human amyloid precursor-like protein 1-cDNA cloning, ectopic
RT expression in COS-7 cells and identification of soluble forms in the
RT cerebrospinal fluid.";
RL Eur. J. Biochem. 250:354-363(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98180887; PubMed=9521588;
RA Lenkkeri U., Kestila M., Lander J., McCready P., Adamson A.,
RA Olsen A., Tryggvason K.;
RT "Structure of the human amyloid-precursor-like protein gene APLP1 at
RT 19q13.1.";
RL Hum. Genet. 102:192-196(1998).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. C-TERMINALLY
CC PROCESSED IN THE GOLGI COMPLEX AND IS THEN SECRETED.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN BRAIN, PARTICULARLY
CC IN THE CEREBRAL CORTEX POSTSYNAPTIC DENSITY.
CC -1- PTM: N- AND O-GLYCOSYLATED.
CC -1- SIMILARITY: BELONGS TO THE APP FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U48437; AAB96331.1; -.
DR EMBL; AD000864; AAB50173.1; -.
DR MIM; 104775; -.
DR InterPro; IPR001868; A4_APP.
DR Pfam; PF02177; A4_EXTRA; 1.
DR PRINTS; PR00203; AMYLOIDA4.
DR SMART; SM00006; A4_EXTRA; 1.
DR PROSITE; PS00319; A4_EXTRA; 1.
DR PROSITE; PS00320; A4_INTRA; 1.
KW Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 21
FT CHAIN 22 650 AMYLOID-LIKE PROTEIN 1.
FT DOMAIN 22 580 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 581 603 POTENTIAL.
FT DOMAIN 604 650 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 640 643 CLATHRIN-BINDING (POTENTIAL).
FT DOMAIN 241 247 POLY-GLU.
FT DOMAIN 264 268 POLY-GLU.
FT CARBOHYD 337 337 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 461 461 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 551 551 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 48 48 P -> A (IN REF. 2).
SQ SEQUENCE 650 AA; 72202 MW; 121A034B708C67CA CRC64;

Query Match 5.48; Score 82.5; DB 1; Length 650;
Best Local Similarity 20.28; Pred. No. 23;
Matches 53; Conservative 25; Mismatches 79; Indels 105; Gaps 12;

QY 53 CAADHRLRLYLSSSLFPMDLIA-----GDDTGAIVITH--LAPQLLEAG 94
DB 84 CLRDQPVRLCYCRQYLPQLQIARVEQATQATPMERWCGSGSRSCAHPHQVVPFRCLPG 143
QY 95 HAFTDAGVRLMLHPQEQFLVNSDRPVESSVR-----AMSAHARVMDGLGLARTPNW 147
DB 144 EFVSEA-----LLVPEGCRFLHQERMDQCESSTRRHRQEAQACSSQGLILHSG----- 192
QY 148 LLLHGGKGGRGAE-LAALIPDLPPVRLRLGLENDERAYSPAELLFICEATGTPPLVFDA 206

```

```

Db 193 MLLPCGSDRFRGVEYVCCPPCTPDPSGTAVG----- 224
QY 207 HHVVVDKLPDOEDFSVREWVLRARATWQPPWQVHLSNGIEGQDQRHSHLIADFSA 266
Db 225 -----DPSRSTRW-----PP-----GSRVEGADEDEEE-----SFPQP 252
QY 267 ----YADVPQIEVEAKGKEEI 284
Db 253 VDDYFVEPQAE-----EETV 270

RESULT 15
DP3A_ECOLI STANDARD; PRT; 1160 AA.
AC P10443;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA POLYMERASE III ALPHA SUBUNIT (EC 2.7.7.7).
GN DNAE OR POLC OR B0184.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88058791; PubMed=3316192;
RA Tomaszewicz H.G., McHenry C.S.;
RT "Sequence analysis of the Escherichia coli dnaE gene.";
RL J. Bacteriol. 169:5735-5744(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MC1555;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / W3110;
RX Takemoto K., Mori H., Murayama N., Kataoka K., Yano M., Itoh T.,
RA Yamamoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R., Ichihara S.,
RA Mizuno T., Makino K., Nakata A., Yura T., Sampei G., Mizobuchi K.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of the
RT 4.0 - 6.0 min (189,987 - 281,416bp) region.";
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA Lashkari D., Lew H., Lin H., Namath A., Oefner P., Roberts D.,
RA Davis R.W.;
RT Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1070-1160 FROM N.A.
RX MEDLINE=93123150; PubMed=7678242;
RA Li S.J., Cronan J.E. Jr.;
RT "Growth rate regulation of Escherichia coli acetyl coenzyme A
RT carboxylase, which catalyzes the first committed step of lipid
RT biosynthesis.";
RL J. Bacteriol. 175:332-340(1993).
RN [6]
RP SEQUENCE OF 1137-1160 FROM N.A.
RX STRAIN-K12 / W3110;
RA Yamamoto Y.;
RT Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [7]
RP REVIEW.
RX MEDLINE=92246902; PubMed=1575709;
RA O'Donnell M.;

```

QY 203 V-----FDAH--IHVVHD--KLPDQEDPSVREWWLRARATWQPPEWQVVHLSN 24

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:51:25 ; Search time 57.72 Seconds
(without alignments)
114.622 Million cell updates/sec

Title: US-09-724-296-39
Perfect score: 1527
Sequence: 1 QLGLVCLTGVPEVRETRVT.....VEAKGKEAIAALRLMAPFK 294

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	94.5	6.2	270	5	PCT-US93-03035-2
2	90	5.9	1421	3	US-09-335-409-2
3	88	5.8	537	2	US-08-633-879C-2
4	85.5	5.6	634	1	US-08-339-152A-17
5	85.5	5.6	653	1	US-08-339-152A-16
6	85.5	5.6	653	2	US-08-007-999B-3
7	85.5	5.6	653	2	US-08-689-276A-3
8	84	5.5	403	4	US-08-861-774E-27
9	83	5.4	1891	2	US-08-804-227C-12
10	83	5.4	1891	2	US-08-804-198-6
11	82.5	5.4	3567	2	US-07-642-734C-4
12	82.5	5.4	3567	3	US-08-439-009A-4
13	81.5	5.3	5588	4	US-09-036-987A-6
14	81.5	5.3	5588	4	US-09-370-700-6
15	81	5.3	607	2	US-08-472-534-5
16	81	5.3	5215	4	US-09-105-537-2
17	80.5	5.3	394	6	5290690-2
18	80	5.2	841	4	US-09-413-814-107
19	80	5.2	7257	3	US-09-335-409-5
20	79.5	5.2	2004	1	US-08-375-709-15
21	79.5	5.2	2004	1	US-08-752-929-15
22	79.5	5.2	2004	4	US-09-090-793-9
23	79	5.2	1841	2	US-08-804-227C-6
24	79	5.2	3816	4	US-09-428-517-3
25	78.5	5.1	394	1	US-07-637-870-1
26	78.5	5.1	394	1	US-07-637-399-1
27	78.5	5.1	394	1	US-07-640-476-5

28	78.5	5.1	394	1	US-08-112-703-1	Sequence 1, Appli
29	78.5	5.1	3491	2	US-07-642-734C-2	Sequence 2, Appli
30	78.5	5.1	3491	3	US-08-439-009A-2	Sequence 2, Appli
31	78	5.1	3519	4	US-09-428-517-4	Sequence 4, Appli
32	77	5.0	587	4	US-08-931-608A-4	Sequence 4, Appli
33	77	5.0	719	4	US-08-975-762-59	Sequence 59, Appl
34	77	5.0	719	4	US-09-295-028-59	Sequence 59, Appl
35	77	5.0	719	4	US-09-106-582-59	Sequence 59, Appl
36	77	5.0	2890	4	US-09-413-814-67	Sequence 67, Appl
37	77	5.0	3798	3	US-09-335-409-6	Sequence 6, Appli
38	77	5.0	4472	2	US-08-804-227C-2	Sequence 2, Appli
39	76	5.0	1088	4	US-09-082-059-2	Sequence 97, Appl
40	76	5.0	1135	2	US-08-469-537A-97	Sequence 3, Appli
41	76	5.0	1864	2	US-08-804-227C-3	Sequence 42, Appl
42	76	5.0	2539	4	US-09-413-814-42	Sequence 8, Appli
43	76	5.0	4550	2	US-08-804-227C-8	Sequence 2, Appli
44	76	5.0	4550	2	US-08-804-198-2	Sequence 2, Appli
45	75.5	4.9	2289	3	US-09-051-019-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
PCT-US93-03035-2
; Sequence 2, Application PC/TUS9303035
; GENERAL INFORMATION:
; APPLICANT: ABBOTT LABORATORIES
; TITLE OF INVENTION: PURIFIED THERMOSTABLE ENDONUCLEASE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES
; STREET: D-377 APGD, ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: ILLINOIS
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03035
; FILING DATE: 19930330
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/869,306
; FILING DATE: 16-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/860,702
; FILING DATE: 31-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BRAINARD, THOMAS D
; REGISTRATION NUMBER: 32,459
; REFERENCE/DOCKET NUMBER: 5145.PC.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-4884
; TELEFAX: 708-937-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 270 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US93-03035-2

Query Match 6.2%; Score 94.5; DB 5; Length 270;
Best Local Similarity 27.5%; Pred. No. 0.017;
Matches 70; Conservative 29; Mismatches 67; Indels 89; Gaps 17;

QY 22 RYRALSAPAREAKLLDLYSSNKTIRGADYCAA--HDRLRYLSS--SLF--PMLDLA 75
 Db 42 RPRALSPAEEA-----FRALREASGGLPAVIAHASYLVNLCAGELMEKSVASLAD 92
 QY 76 DDTGAVALTHLAPQLLEAGHAFTDAGR--LLMHPEQFIVLNSDRPE--VRESSVRASAH 133
 Db 93 DLEKAALL-----GVEVVVHP-----GSRPERVKEGALKALRL-- 127
 QY 134 RVMGIGLGARTPNWLLLLH-----GGK--GGRGAELAAALIPDLPVRLRLGLENDERAYS 187
 Db 128 -----AGVRSRP--VLLVENTAGGKVGARPEELAWLVAADTPLOV----- 166
 QY 188 PAELLPICATGTPLVPDAHHVHVDKLPDQEDPSVREHVLBARATWOPPEWQVHLSNG 247
 Db 167 ---CLDTCHA-----YAGYDVAEDPL-----GVLDALDRAVGLRVP-----VHLNDS 208
 QY 248 IEGPQDR--RHSHLI 260
 Db 209 VGLGSRVDHHAHL 223

RESULT 2
 US-09-335-409-2
 ; Sequence 2, Application US/09335409
 ; Patent No. 6121029
 ; GENERAL INFORMATION:
 ; APPLICANT: Schupp, Thomas
 ; APPLICANT: Ligon, James
 ; APPLICANT: Molnar, Istvan
 ; APPLICANT: Zirkle, Ross
 ; APPLICANT: Cyr, Devon
 ; APPLICANT: Goerlach, Joern
 ; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
 ; FILE REFERENCE: 4-30582A
 ; CURRENT APPLICATION NUMBER: US/09/335,409
 ; CURRENT FILING DATE: 1999-06-17
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 2
 ; LENGTH: 1421
 ; TYPE: PRT
 ; ORGANISM: Sorangium cellulosum
 US-09-335-409-2

Query Match 5.9%; Score 90; DB 3; Length 1421;
 Best Local Similarity 21.5%; Pred. No. 0.78;
 Matches 78; Conservative 37; Mismatches 122; Indels 126; Gaps 19;

QY 5 VCLAVGPEVRFTVLSRYRALSAPAREAKLLDLY-----SSNKTIRGADYCA 54
 Db 129 VFIGIP-----SEYEALPQATASAEIDAHAHGLTMTSPVSGAGRISYALGLRGPV 179
 QY 55 AHDLRLYRLSSLPMLDLADGDDTGAVALTHLAPQLLEAGHAFT--DAGVRLMHPEQFI 112
 Db 180 AVDT---AYSSL-----VAVHLACQSRSGECSTALAGGVSLSLSPSTLV 222
 QY 113 VLNSDRPREVRESSVRASAHARVMDGLGLARTPNWLL--LHGCK--GGR----- 158
 Db 223 WLSKTRALARADGRCKAFSAEA---DGEGRGCGCAVVVLKRLSGARADGDRILAVIRGSAI 279
 QY 159 ---GAELAAALPD-----LPDPVRLR-----L 177
 Db 280 NHDGASSGLTVPNGSSQEIVLKRALADAGCAASVGVVEAHGTGTTGLGDPETQALNAVY 339
 QY 178 GLENDERAYSAPAEILLPICEATGTP-----LVFDAHHVHVHDKLPDQE--DPSV-- 223
 Db 340 GLGRD--VATPLLGSVKTNLGHPEYASGTTGLKLVLSLQHGQIPAHLLHAQALNPRISW 397
 QY 224 ---REWLRLARATWOPPEWQVH-----LSNIEGQPDQRHSHLI--ADEPSPAYADVPIEV 275
 Db 398 GDLRLTVTRATPW--PDWNTPRRAGVSSFGMSG-----TNAVHVLEEPAPATCTPPAPER 451

QY 276 EAK 278
 Db 452 PAE 454

RESULT 3
 US-08-633-879C-2
 ; Sequence 2, Application US/08633879C
 ; Patent No. 5928922
 ; GENERAL INFORMATION:
 ; APPLICANT: Kivirikko, Kari I.
 ; APPLICANT: Pihlajaniemi, Taina
 ; APPLICANT: Helaakoski, Tarja I.
 ; APPLICANT: Annunen, Pia P.
 ; APPLICANT: Nissi, Riitta K.
 ; APPLICANT: No. 5928922elainen, Minna K.
 ; TITLE OF INVENTION: 2 SUBUNIT OF PROLYL-4-HYDROXYLASE
 ; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING SUCH SUBUNIT AND
 ; TITLE OF INVENTION: METHODS FOR PRODUCING THE SAME
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds, LLP
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: NY
 ; COUNTRY: USA
 ; ZIP: 10036-2811
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: Windows
 ; SOFTWARE: FastSeq for Windows Version 2.0b
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/633,879C
 ; FILING DATE: 10-APR-1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Abrams, Samuel B
 ; REGISTRATION NUMBER: 30,605
 ; REFERENCE/DOCKET NUMBER: 8389-0041-999
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650-493-4935
 ; TELEFAX: 650-493-5556
 ; TELEX: 66141 PENNIE
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 537 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; FRAGMENT TYPE: internal
 US-08-633-879C-2

Query Match 5.8%; Score 88; DB 2; Length 537;
 Best Local Similarity 22.2%; Pred. No. 0.28;
 Matches 90; Conservative 39; Mismatches 116; Indels 160; Gaps 22;

QY 32 EAKLLDLYSSNIK-----TLKGAAD---YCAAHDIRLYRISSSL---FPM-L-DLGGDDTCA 80
 Db 54 EAKLAKIKSWASKMEALTSKSAADPEGY-LAHPVNAVYKLVKRLNTDWPALGDLVLQDASA 112
 QY 81 AVLTHLAPOLL-----EAGHAFTDAGVRLMHPEQFIVLNSD---RPEVRESSVRAMS 130
 Db 113 GFVANLSVGRQFPTDEDESGAA-----RALMRLQDTYKLDPDITISRGELPTKYQAML 166
 QY 131 AHARVMDGLGLARTPNW-----LILLHGKGG----- 157
 Db 167 S----VDDCFGLGRSATNEGDYHTVLMMEQVLKOLDAGEAATVTKSLVDLYLSYAVFQLG 223

QY 158 ---RGAELAAALIPDLPVRLRLG-----LENDERAYS-----P 188
DB 224 DLHRAVELTRLLSL-DFSHRAGNRLYFERLLEERGGKSLSNQTDAGLATQENLYERP 282
QY 189 AELLP-----ICEATGTPL-----VFDAHH-----HVV- 211
DB 283 TDYLPEDVYSLCGEGVKLTTPRQKLCFCRYHHGNRVPLLIAPTKEEDWDSPIHVR 342
QY 212 -HDKLPDOEDPSVREWV--LRARATWQPPPEWQV-----VHLSNGIEGQD-----R 254
DB 343 YDVSDEEIERKEIAKPKIARATVDPKGTGVLTVASRYVSKSWLEDDDPVVARVNR 402
QY 255 RHSHL-----IADPPSAVADVPQIEVEAKKEEAIAL 287
DB 403 RMOHTGLTVKTAELLOVANYGMGQIEPHDFSRSDDEDAFKRL 447

RESULT 4
US-08-339-152A-17
: Sequence 17, Application US/08339152A
: Patent No. 5643726
: GENERAL INFORMATION:
: APPLICANT: Tanzi, Rudolph E.
: APPLICANT: Kovacs, Dora M.
: TITLE OF INVENTION: Methods For Modulating Transcription
: TITLE OF INVENTION: From The Amyloid -Protein Precursor (APP) Promoter
: NUMBER OF SEQUENCES: 33
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
: STREET: 1100 New York Ave., NW, Suite 600
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/339,152A
: FILING DATE: 10-NOV-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Steffe, Eric K.
: REGISTRATION NUMBER: 36,688
: REFERENCE/DOCKET NUMBER: 0609.4120000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-371-2600
: TELEFAX: 202-371-2540
: TELEX:
: INFORMATION FOR SEQ ID NO: 17:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 634 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
US-08-339-152A-17

Query Match 5.6%; Score 85.5; DB 1; Length 634;
Best Local Similarity 21.2%; Pred. No. 0.69;
Matches 55; Conservative 22; Mismatches 83; Indels 99; Gaps 10;

QY 53 CAAHDIRLYRLSSSLFPMPLDLA-----GDDTGAVALTH--LAPQLLEAG 94
DB 63 CLLDQPVRLCYCRQMPYELHIAVEQAQAIPMERWCGTGRSCRAHPHHEVVPFHCLPG 122
QY 95 HAFTDAGVRLLMHPQFIVLNSDRPREVRESSVR-----AMSAHARVMDGLGARTPNW 147
DB 123 EFWSEA----LLVPEGCRFLHQERMDQCESSTRRHQEAQACSSQGLILHSG----- 171
QY 148 LLLHGGKGRGAELAAALIPDLPVRLRLGLENDERAYSAPPELLPICEATGTPLVFDH 207

DB 172 MLLPCGSDRFRGVEYVCCPP-----PATPNPSGMAAG----- 203
QY 208 HHVVHDKLPDOEDPSVREWVLRARATWQPPPEWQVHLSNGIEGPQDRIHSHLIADFPSPA- 266
DB 204 -----DPSTRSWPLGGRA-----EGGEDEEE---VESFPQPV 232
QY 267 ---YADVPOJFEVAKKEE 282
DB 233 DDFVEPFA-----EEEEEEEE 251

RESULT 5
US-08-339-152A-16
: Sequence 16, Application US/08339152A
: Patent No. 5643726
: GENERAL INFORMATION:
: APPLICANT: Tanzi, Rudolph E.
: APPLICANT: Kovacs, Dora M.
: TITLE OF INVENTION: Methods For Modulating Transcription
: TITLE OF INVENTION: From The Amyloid -Protein Precursor (APP) Promoter
: NUMBER OF SEQUENCES: 33
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
: STREET: 1100 New York Ave., NW, Suite 600
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/339,152A
: FILING DATE: 10-NOV-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Steffe, Eric K.
: REGISTRATION NUMBER: 36,688
: REFERENCE/DOCKET NUMBER: 0609.4120000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-371-2600
: TELEFAX: 202-371-2540
: TELEX:
: INFORMATION FOR SEQ ID NO: 16:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 653 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-339-152A-16

Query Match 5.6%; Score 85.5; DB 1; Length 653;
Best Local Similarity 21.2%; Pred. No. 0.73;
Matches 55; Conservative 22; Mismatches 83; Indels 99; Gaps 10;

QY 53 CAAHDIRLYRLSSSLFPMPLDLA-----GDDTGAVALTH--LAPQLLEAG 94
DB 83 CLLDQPVRLCYCRQMPYELHIAVEQAQAIPMERWCGTGRSCRAHPHHEVVPFHCLPG 142
QY 95 HAFTDAGVRLLMHPQFIVLNSDRPREVRESSVR-----AMSAHARVMDGLGARTPNW 147
DB 143 EFWSEA----LLVPEGCRFLHQERMDQCESSTRRHQEAQACSSQGLILHSG----- 191
QY 148 LLLHGGKGRGAELAAALIPDLPVRLRLGLENDERAYSAPPELLPICEATGTPLVFDH 207
DB 192 MLLPCGSDRFRGVEYVCCPP-----PATPNPSGMAAG----- 223
QY 208 HHVVHDKLPDOEDPSVREWVLRARATWQPPPEWQVHLSNGIEGPQDRIHSHLIADFPSPA- 266

```

Db 224 -----DPSTRSWPLGRRA-----EGDEDEE---VESFPQPV 252

Qy 267 ---YADVPQIEVEAKGKEE 282
      : | | | : | |
Db 253 DDYFVEPPQAEHEEEEEE 271

RESULT 6
US-08-007-999B-3
: Sequence 3, Application US/08007999B
: Patent No. 5851787
: GENERAL INFORMATION:
: APPLICANT: wasco, Wilma
: APPLICANT: Bupp, Keith
: APPLICANT: Magendantz, Margaret
: APPLICANT: Tanzi, Rudolph
: APPLICANT: Solomon, Frank
: TITLE OF INVENTION: AMYLOID PRECURSOR-LIKE PROTEIN AND USES
: TITLE OF INVENTION: THEREOF
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sterne, Kessler, Goldstein & Fox, P.L.L.C.
: STREET: 1100 New York Ave., NW
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/007,999B
: FILING DATE: 21-JAN-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/872,642
: FILING DATE: 20-APR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/930,022
: FILING DATE: 17-AUG-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Townsend, G. Kevin
: REGISTRATION NUMBER: 34,033
: REFERENCE/DOCKET NUMBER: 0609.3520002/JAG/GKT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202)371-2571
: TELEFAX: (202)371-2540
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 653 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-007-999B-3

```

Query Match 5.6%; Score 85.5; DB 2; Length 653;
Best Local Similarity 21.2%; Pred. No. 0.73;
Matches 55; Conservative 22; Mismatches 83; Indels 99; Gaps 10;

[illegible]

```

Qy  208 HHVVHDKLPDQEPFVREKWLRLARATWQPPQVHLSNGIEGQPRRHHLIADFPSPA - 266
Db  224 -----DPSTRSWPLGGRA-----EGGEDEEE---VESFPQP 252
Qy  267 --YADVPQIEVEAKGEE 282
Db  253 DDYFVEPQAEHEEEEF 271
RESULT 7
US-08-689-276A-3
; Sequence 3, Application US/08689276A
; Patent No. 5891591
; GENERAL INFORMATION:
; APPLICANT: wasco, Wilma
; APPLICANT: Bupp, Keith
; APPLICANT: Magendantz, Margaret
; APPLICANT: Tanzi, Rudolph
; APPLICANT: Solomon, Frank
; TITLE OF INVENTION: AMYLOID PRECURSOR-LIKE PROTEIN AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Ave., NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/689,276A
; FILING DATE: 06-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/007,999
; FILING DATE: 21-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,642
; FILING DATE: 20-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,022
; FILING DATE: 17-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: JORGE A. GOLDSTEIN
; REGISTRATION NUMBER: 29,021
; REFERENCE/DOCKET NUMBER: 0609.9520003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2571
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 653 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-689-276A-3

```

Query Match 5.6%; Score 85.5; DB 2; Length 653;
 Best Local Similarity 21.2%; Pred. No. 0.73;
 Matches 55: Conservative 22; Mismatches 83; Indels 99; Gaps 10;

QY 53 CAADHTIRYLKSSLSFLPMDLA-----GDGTGA AVLTH- LAPQLLEAG 94
| : : : :
Db 83 CLDPQRVLEYCYMPCHIAIEVQAQAIPMERWCGGTRSGCAHPHHEVPFCLPG 142
| : : : :
QY 95 HAFTDAGVRLIMHPEOFIVLNSDRPFRESVR-----ANSAIKRVMDGLGLARTPNW 147
| : : : :
Db 143 EFWSEA----LLVPECKRFUHOERMDOCSSTRHHOEAECSSGLI LHGSG----- 191

;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/439,009A
;; FILING DATE: 11-MAY-1995
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Casuto, Dianne
;; REGISTRATION NUMBER: 40,943
;; REFERENCE/DOCKET NUMBER: 4952.US.D1
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 847-938-3137
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3567 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-439-009A-4

Query Match 5.4%; Score 82.5; DB 3; Length 3567;
Best Local Similarity 23.8%; Pred. No. 24;
Matches 64; Conservative 28; Mismatches 94; Indels 83; Gaps 13;

Qy 18 VTLGRALSP-----AREAKLLDLYSSNKTGRGAADYCAAHDIRLYRLSSLPMLDL 73
Db 1847 VVLAHRGMLPRSLHADSPLHIDWESGAVEVLEEVPPWAGERPRRAGVSS-----FGV 1901
Qy 74 AGDDTGRAVLTHLAPOLLEAGHAFTDAGVRLLMHPEQFIVLNSDRPE-VRESSVRAMSAH 132
Db 1902 SG--TNAHVIVVEAPAEQEA--ARTERG-----PLPFVL--SGRSEAVVAAQARALAEH 1949
Qy 133 ARVMDGLGLARTPNWL-----LLHGGKGGGGAELAAALIPOLPDVPRVRLGLEN 181
Db 1950 LRDTPGLGLTDAWTLATGRADFVRAAVLGDGDRAGVCAELDLAEGRP----- 1998
Qy 182 DERAYSPAELLPICEATGTP-LVFDAAHHVVDKLPDQEDPSVREWVLARATWQPE-- 238
Db 1999 -----SADAVAPVTSAPRKPVLP-----PGQG-----AQWVGMDRLLESSEVF 2038
Qy 239 -----WQVHLSNGIEGP 251
Db 2039 AESMSRCAELSPHTDMKLLDVVRGDDGP 2067

RESULT 13

US-09-036-987A-6
; Sequence 6, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow Agrosciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/036,987A
;; FILING DATE: 09-MAR-1998
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Stuart, Donald R.
;; REGISTRATION NUMBER: 28,479
;; REFERENCE/DOCKET NUMBER: 50,608
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (317)337-4816
;; TELEFAX: (317)337-4847
;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 5588 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-09-036-987A-6

Query Match 5.3%; Score 81.5; DB 4; Length 5588;
Best Local Similarity 21.8%; Pred. No. 64;
Matches 62; Conservative 31; Mismatches 103; Indels 89; Gaps 12;

Qy 22 RYALSPAEREAKLLDLYSSNKTGRG-----AADYCAAHDIRLYRLSS 65
Db 5204 RLAGLSEGERQQVQLQVRADIAVILGHGRSSDVDTIEKPLAELGFDLSIAELR-NRLAT 5262
Qy 66 SLFPML--DLAGD-DTCAAVLTHLAPQLLEAGHAFTDAGVRLLMHPEQFIVLNSDRPEVR 122
Db 5263 ATGLRLPATLAFDHGTAAALAOHVCAQL--GTATAPAPRR-----TDNDATPEV 5310
Qy 123 ESSVRAMSAHARVMDGLGLARTPNWLHLLHGGKGGGGAELAAALIPOLPDVPRVRLGLEND 182
Db 5311 RSLFQQAAYAGRLDGMVLKVAQAQLRPVFGSPGE-----LESILPKPVQLSRG---- 5358
Qy 183 ERAYSPAELLPICEATGTP-LVFDAAHHVVDKLPDQEDPSVREWVLARATWQPEWQVV 242
Db 5359 -----PEELALVC-----MPALIG-----MPPAQQYA 5380
Qy 243 HLSNGIEGPQDRRHSHLIADFPSAYADVP---QIEVEAKGKEAI 284
Db 5381 RTAAGFRDVRDVS-----VIPMPGFIAEPLPSAIEVAVRTQAEAV 5421

RESULT 14

US-09-370-700-6
; Sequence 6, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT APPLICATION NUMBER: US/09/370,700
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 6
; LENGTH: 5588
; TYPE: PRT
; ORGANISM: Saccharopolyspora spinosa
US-09-370-700-6

Query Match 5.3%; Score 81; DB 2; Length 607;
Best Local Similarity 21.0%; Pred. No. 2.1;

Search completed: January 15, 2002, 13:51:29
Job time: 358 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 14:07:20 ; Search time 45.07 Seconds
(without alignments)
471.835 Million cell updates/sec

Title: US-09-724-296-38
Perfect score: 3055
Sequence: 1 MGTTLGSLSLGRGAAPTV.....HKLNKYHDWLWENHEKLSLS 580

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	208.5	6.8	486	1 MEC2_HUMAN	P51608 homo sapien
2	201	6.6	492	1 MEC2_RAT	Q00566 rattus norv
3	198.5	6.5	484	1 MEC2_MOUSE	Q92246 mus musculu
4	145	4.7	1301	1 SAC3_YEAST	P46674 saccharomyc
5	145	4.7	3418	1 BRC2_HUMAN	P51587 homo sapien
6	138.5	4.5	500	1 GAR2_SCHPO	P41891 schizosacch
7	137	4.5	560	1 YJK5_YEAST	P42948 saccharomyc
8	136.5	4.5	598	1 CYL1_HUMAN	P35663 homo sapien
9	133	4.4	1447	1 BUD4_YEAST	P47136 saccharomyc
10	131.5	4.3	997	1 BTR1_SCHPO	O14064 schizosacch
11	122	4.0	3924	1 ANK2_HUMAN	Q01484 homo sapien
12	120.5	3.9	1147	1 AC15_HUMAN	P35251 homo sapien
13	120.5	3.9	1233	1 YF16_YEAST	P43597 saccharomyc
14	120	3.9	1131	1 AC15_MOUSE	P35601 mus musculu
15	120	3.9	1658	1 YMK7_YEAST	Q03661 saccharomyc
16	119	3.9	886	1 ORC1_KLULA	P54788 kluyveromyc
17	118	3.9	704	1 NP14_RAT	P41777 rattus norv
18	118	3.9	3329	1 MEC2_MOUSE	P97929 mus musculu
19	117	3.8	1273	1 YAR2_SCHPO	Q10135 schizosacch
20	117	3.8	1850	1 VIT2_CHICK	P02845 gallus gall
21	116.5	3.8	633	1 MLH1_TETTH	P40631 tetrahymena
22	116	3.8	1290	1 XPCX_XENLA	P50532 xenopus lae
23	116	3.8	3256	1 K167_HUMAN	P46013 homo sapien
24	115.5	3.8	2156	1 ORP1_HUMAN	P56715 homo sapien
25	114.5	3.7	970	1 PSU1_YEAST	P53550 saccharomyc
26	114.5	3.7	1213	1 FMN_CHICK	Q05858 gallus gall
27	114	3.7	783	1 YAYB_SCHPO	O10218 schizosacch
28	114	3.7	805	1 TACL1_HUMAN	O75410 homo sapien
29	114	3.7	1539	1 Y373_HUMAN	O15078 homo sapien
30	113.5	3.7	1164	1 BAG_STRAG	P27951 streptococc
31	113	3.7	482	1 YSR2_CAEEL	Q09950 caenorhabdi
32	113	3.7	727	1 PEC1_MOUSE	Q08481 mus musculu
33	112.5	3.7	676	1 HS7C_TRYBB	P20030 trypanosoma

ALIGNMENTS

```

RESULT 1
MEC2_HUMAN
ID MEC2_HUMAN STANDARD; PRT; 486 AA.
AC P51608; O15233;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE METHYL-CPG-BINDING PROTEIN 2 (MECP-2 PROTEIN) (MECP2).
GN MECP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kudo S., Fukuda M.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Thiesen J., Straetling W.H.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE-97130625; PubMed-8976388;
RA Vilain A., Apiou F., Vogt N., Dutrillaux B., Malfroy B.;
RT "Assignment of the gene for methyl-CpG-binding protein 2 (MECP2) to
RT human chromosome band Xq28 by in situ hybridization.";
RN [4]
RP Cytogenet. Cell Genet. 74:293-294(1996).
RN [5]
RP SEQUENCE FROM N.A.
RA Reichwald K., Rosenthal A., Kioschis P., Platzer M.;
RT "Mapping and sequence analysis of the human MECP2 locus.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 10-486 FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE-96327611; PubMed-8672133;
RA D'Esposito M., Quaderi N.A., Ciccocioppa A., Bruni P., Esposito T.,
RA D'Esposito M., Brown S.D.M.;
RT "Isolation, physical mapping, and Northern analysis of the X-linked
RT human gene encoding methyl Cpg-binding protein, MECP2.";
RL Mamm. Genome 7:533-535(1996).
RN [7]
RP SEQUENCE OF 10-486 FROM N.A.
RA Reichwald K., Bauer D., Brenner V., Drescher B., Coy J.,
RA Kioschis P., Korn B., Nyakatura G., Platzer M., Poustka A.,
RA Sandoval N., Rosenthal A.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [8]
RP VARIANTS RTT TRP-106; CYS-133; SER-155 AND MET-158.
RX MEDLINE-99438392; PubMed-10508514;
RA Amir R.E., Van den Veyver I.B., Wan M., Tran C.Q., Francke U.,
RA Zoghbi H.Y.;
RT "Rett syndrome is caused by mutations in X-linked MECP2, encoding
RT methyl-CpG-binding protein 2.";
```

009750 schizosacch
 P30414 homo sapien
 Q01448 saccharomyc
 P49452 mus musculu
 P35201 saccharomyc
 Q24595 drosophila
 P20357 mus musculu
 P35662 bos taurus
 P39526 saccharomyc
 P17141 mus musculu
 Q13435 homo sapien
 P20676 saccharomyc

RL Nat. Genet. 23:185-188(1999).
 RN [8]
 RP VARIANT RTT VAL-140.
 RX MEDLINE=20465115; PubMed=11007980;
 RA Orrico A., Lam C., Galli L., Dotti M.T., Hayek G., Tong S.F.,
 RA Poon P.M., Zappella M., Federico A., Sorrentino V.;
 RT "MECP2 mutation in male patients with non-specific X-linked mental
 RT retardation.";
 RL FEBS Lett. 481:285-288(2000).
 RN [9]
 RP VARIANTS RTT W-106; P-124; C-13; C-134; R-152; M-158 AND C-306.
 RX MEDLINE=20439334; PubMed=10991688;
 RA Obata K., Matsui T., Yamashita Y., Fukuda T., Kuwajima K.,
 RA Horiuchi I., Nagamitsu S., Iwanaga R., Kimura A., Omori I., Endo S.,
 RA Mori K., Kondo I.;
 RT "Mutation analysis of the methyl-CpG binding protein 2 gene (MECP2) in
 RT patients with Rett syndrome.";
 RL J. Med. Genet. 37:608-610(2000).
 RN [10]
 RP VARIANTS RTT R-101; W-106; M-158 AND C-306, AND VARIANT K-397.
 RX MEDLINE=20439335; PubMed=10991689;
 RA Hampson K., Woods C.G., Latif F., Webb T.;
 RT "Mutations in the MECP2 gene in a cohort of girls with Rett
 RT syndrome.";
 RL J. Med. Genet. 37:610-612(2000).
 CC -!- FUNCTION: CHROMOSOMAL PROTEIN THAT BINDS TO METHYLATED DNA. IT CAN
 CC BIND SPECIFICALLY TO A SINGLE METHYL-CPG PAIR. IT IS NOT
 CC INFLUENCED BY SEQUENCES FLANKING THE METHYL-CPGs. MEDIATES
 CC TRANSCRIPTIONAL REPRESSION THROUGH INTERACTION WITH HISTONE
 CC DEACETYLASE AND THE COREPRESSOR SIN3A.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR. COLOCALIZED WITH METHYL-CPG IN THE
 CC GENOME.
 CC -!- TISSUE SPECIFICITY: PRESENT IN ALL ADULT SOMATIC TISSUES TESTED.
 CC -!- DISEASE: DEFECTS IN MECP2 ARE THE CAUSE OF RETT SYNDROME (RTT), AN
 CC X-LINKED DOMINANT DISEASE. RTT IS A PROGRESSIVE NEUROLOGIC
 CC DEVELOPMENTAL DISORDER AND ONE OF THE MOST COMMON CAUSES OF MENTAL
 CC RETARDATION IN FEMALES. PATIENTS APPEAR TO DEVELOP NORMALLY UNTIL
 CC 6 TO 18 MONTHS OF AGE, THEN GRADUALLY LOOSE SPEECH AND PURPOSEFUL
 CC HAND MOVEMENTS AND DEVELOP MICROCEPHALY, SEIZURES, AUTISM, ATAXIA,
 CC INTERMITTENT HYPERVENTILATION, AND STEREOTYPIC HAND MOVEMENTS.
 CC AFTER INITIAL REGRESSION, THE CONDITION STABILIZES AND PATIENTS
 CC USUALLY SURVIVE INTO ADULTHOOD.
 CC -!- SIMILARITY: CONTAINS (R)GRP(K) MOTIFS AND SPKK MOTIFS THAT HAVE
 CC BEEN FOUND TO BIND TO MINOR GROOVE OF AT-RICH DNA.
 CC -!- SIMILARITY: CONTAINS 1 METHYL-BINDING DOMAIN (MBD).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; L37298; AAC32737.1; -;
 DR EMBL; Y12643; CAA73190.1; -;
 DR EMBL; X99686; CAA68001.1; -;
 DR EMBL; AF030876; AAC08757.1; -;
 DR EMBL; AF031078; AAC08758.1; -;
 DR EMBL; X89430; CAA61599.1; -;
 DR EMBL; X94628; CAA64331.1; -;
 DR MIM; 300005; -;
 DR MIM; 312750; -;
 DR InterPro; IPR000637; AT hook.
 DR InterPro; IPR001739; MBD.
 DR Pfam; PF01429; MBD; 1.
 DR SMART; SM00384; AT_hook; 1.
 DR SMART; SM00391; MBD; 1.
 KW Transcription regulation; Repressor; DNA-binding; Nuclear protein;
 KW Disease mutation; Polymorphism.
 FT DOMAIN 96 149 MBD.
 FT DOMAIN 277 283 POLY-ALA.
 FT DOMAIN 366 372 POLY-HIS.

FT DOMAIN 384 393 POLY-PRO.
 FT VARIANT 101 101 P -> R (IN RTT).
 FT 106 106 /FTID=VAR_010276.
 FT 124 124 R -> W (IN RTT).
 FT 133 133 /FTID=VAR_010272.
 FT 134 134 L -> F (IN RTT).
 FT 152 152 /FTID=VAR_010277.
 FT 155 155 R -> C (IN RTT).
 FT 158 158 /FTID=VAR_010273.
 FT 201 201 S -> C (IN RTT).
 FT 290 290 /FTID=VAR_010278.
 FT 306 306 A -> V (IN RTT).
 FT 306 306 /FTID=VAR_010279.
 FT 397 397 P -> R (IN RTT).
 FT 397 397 /FTID=VAR_010280.
 FT 72 72 E -> S (IN RTT).
 FT 290 290 /FTID=VAR_010274.
 FT 486 486 T -> M (IN RTT).
 FT 201 201 /FTID=VAR_010275.
 FT 306 306 A -> V.
 FT 397 397 R -> C (IN RTT).
 FT 397 397 /FTID=VAR_010282.
 FT 72 72 E -> K.
 FT 290 290 /FTID=VAR_010283.
 FT 486 486 PAVP -> RLC (IN REF. 5).
 FT 290 290 E -> G (IN REF. 3).
 SQ SEQUENCE 486 AA; 52440 MW; EB6A33233AEDA566 CRC64;
 Query Match 6.8%; Score 208.5; DB 1; Length 486;
 Best Local Similarity 23.9%; Pred. No. 2.5e-06;
 Matches 104; Conservative 53; Mismatches 149; Indels 129; Gaps 17;
 QY 37 EDVAMELERVCEDEEOMNKRSECNPLLOEPIASAOFG-----ATACT 80
 DB 22 KDRPLKFKVKYKDKKEGKHEPVOPSAHSAEPAEACKAETSEGGSPAPVPEASASP 81
 QY 81 ECRKSV-----PCGWERVVKQRLFOKTAGRFDVYFISQGLKFRSKSLANY 127
 DB 82 KQRSITRDGPMYDDPTLPEGWTRKLQKRSRAGKYDVLINPOGKAFRSKVELIAY 141
 QY 128 LKNGETSLKPEDFDFTVLSKRGIKSRKYDCSMAALTSHLQNSNNSNNMLRTRSKCKKD 187
 DB 142 FEKVGDTSLDNDFDFTV-TGRGSPSR-----RQKPPKK- 175
 QY 188 VFMPSSSSSELQSRGL---SNETSTHLLKDEGVDDVNFVRKVRPKGVILKGIPIK 244
 DB 176 ---PKSPKAPGTGRGRGRPKGSGTTRPKAATSEGVQVK--RVLEKSPGK--LLVKMPF- 226
 QY 245 KTKKGRKSCSGFVQSDS---KRESVCNKADAESEPVAKQSOLDRTVCISDAGACGTEL 300
 DB 227 QTSFGKAECCGGATTSTQWVVKRPGKRRKAEADPAIPKKR-----GRKPGSVV 276
 QY 301 SVTSEENSLVKKERSLSGSCNFCSEQKTSGLINKFCSAKDEHNEKEYEDTFLESEIGT 360
 DB 277 AAAAAEAKKAVKESUR-----SVQETVLPKK-----RKTRET-----V 312
 QY 361 KVEVVERKEHLHTDIL--KRGSEMDNNCSPTKDFGCEKIFQEDTIPRTQIRKRTSLYF 418
 DB 313 STIEVKVVKPLVLTILGKSGKGLKTKCSGPKR-----SKESPCKGR----- 354
 QY 419 SSKYNKEALSPRRKK 433
 DB 355 ----SSASPPKKE 365
 RESULT 2
 MEC2_RAT
 ID MEC2_RAT STANDARD; PRT; 492 AA.
 AC Q00566;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)

20-AUG-2001 (Rel. 40, Last annotation update)
 METHYL-CPG-BINDING PROTEIN 2 (MECP-2 PROTEIN) (MECP2).
 DE MECP2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Brain;
 RX MEDLINE=92298389; PubMed=1606614;
 RA Lewis J.D., Meenan R.R., Henzel W.J., Maurer-Fogy I., Jeppesen P.,
 RA Klein F., Bird A.;
 RA "Purification, sequence, and cellular localization of a novel
 FT chromosomal protein that binds to methylated DNA.";
 RL Cell 69:905-914(1992).
 CC -!- FUNCTION: CHROMOSOMAL PROTEIN THAT BINDS TO METHYLATED DNA. IT CAN
 CC BIND SPECIFICALLY TO A SINGLE METHYL-CPG PAIR. IT IS NOT
 CC INFLUENCED BY SEQUENCES FLANKING THE METHYL-CPGS. MEDIATES
 CC TRANSCRIPTIONAL REPRESSION THROUGH INTERACTION WITH HISTONE
 CC DEACETYLASE AND THE COREPRESSOR SIN3A (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: NUCLEAR. COLOCALIZED WITH METHYL-CPG IN THE
 CC GENOME.
 CC -!- TISSUE SPECIFICITY: PRESENT IN ALL ADULT SOMATIC TISSUES TESTED.
 CC -!- SIMILARITY: CONTAINS (R)GRP(K) MOTIFS AND SPK MOTIFS THAT HAVE
 CC BEEN FOUND TO BIND TO MINOR GROOVE OF AT-RICH DNA.
 CC -!- SIMILARITY: CONTAINS 1 METHYL-BINDING DOMAIN (MBD).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M94064; AAA11584.1; -;
 DR PIR; A41907; A41907.
 DR InterPro; IPR001739; MBD.
 DR Pfam; PF01429; MBD; 1.
 DR SMART; SM00391; MBD; 1.
 KW Transcription regulation; Repressor; DNA-binding; Nuclear protein.
 FT DOMAIN 96 149 MBD.
 FT DOMAIN 366 372 POLY-HIS.
 FT DOMAIN 384 391 POLY-PRO.
 FT DOMAIN 443 451 POLY-THR.
 FT SEQUENCE 492 AA; 53047 MW; A67E705C68BA2D38 CRC64;
 SQ
 Query Match 6.6%; Score 201; DB 1; Length 492;
 Best Local Similarity 22.9%; Pred. No. 7.8e-06;
 Matches 108; Conservative 57; Mismatches 156; Indels 150; Gaps 18;
 QY 34 LRKE-----DVAMELRVGEDEEOMMKRRSECNPLLOEPIASAQFG-----75
 Db 8 LRKEKESDQLQGLKEPLKFKKVKKKEDKKGKLEPLQPSAHHSAPAEAGKAETSES 67
 QY 76 -----ATAGTECRKSY-----PCGWERVVKORLFGKTAGRDVYFISP 113
 Db 68 SGSAPVPEASAPKQRRIIRGPMVDDPTLPEGWTRKLKQKRGSRGAKYDVYLYNP 127
 QY 114 QGLKFRSKSLANYLKHNGTSLKPEDFDFTVLSKRGKISRYKDCSMAALTSHLQNSNN 173
 Db 128 QGKAFRSKVELIAYFEKVGDTSLDPNDFDTV-TGRGSPSR-----167
 QY 174 SNWNILRTSRCKKDVENPPSSSELSQESRGLSNFTSTHLLKDEGVDVNNFKV-RKPK 232
 Db 168 -----RQKPPKK-----PKSPKAPGTGGRGRPKSGTGRPKAAASGEVQVKRVLEKSP 217
 QY 233 GKVTILKGIPIKTKKCRKSCSGFVQSDS-----KRESCVNKAADAESEVQAKSOLDRT 287
 Db 218 GK--LLVKMPFOASPGG--KGGGGATTSAQVMVIRPKGRKRAEADPOAIPKR-----268

QY 288 VCISDAGACGETLSVTSEENSLVKKERSLSSGNSFCSEQKTSGLINKFCSAKDSEHNEK 347
 Db 269 -----CRKPGSVVAAAAEAACKKAVKESSIR-----SVQETVLPKK-----RK 307
 QY 348 YEDTFLESEEGIKVVEVERKEHLHWDIL--KRGSEMDNNCSPTTRKDTGKIFQEDTIP 405
 Db 308 TRET-----VSIEVKVVKVFLVSTLGEKSGKGLKTKCKSPGRK-----SKESP 351
 QY 406 RTQIERRTSLYFESSYKNEALSPRR-----KAFKKWTTPPRSP 444
 Db 352 KGR-----SSASSPPKKEHHHHHHHAESPAPMLPLPPPPP 388
 RESULT 3
 MEC2_MOUSE
 ID MEC2_MOUSE STANDARD; PRT; 484 AA.
 AC Q9Z2D6;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE METHYL-CPG-BINDING PROTEIN 2 (MECP-2 PROTEIN) (MECP2).
 GN MECP2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6;
 RX MEDLINE=98449942; PubMed=9774669;
 RA Hendrich B., Bird A.;
 RT "Identification and characterization of a family of mammalian methyl-
 RT Cpg binding proteins.";
 RL Mol. Cell. Biol. 18:6538-6547(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99299240; PubMed=10369871;
 RA Coy J.F., Sedlacek Z., Baechner D., Delli H., Poustka A.;
 RT "A complex pattern of evolutionary conservation and alternative
 RT polyadenylation within the long 3'-untranslated region of the methyl-
 RT Cpg-binding protein 2 gene (MeCP2) suggests a regulatory role in gene
 RT expression.";
 RL Hum. Mol. Genet. 8:1253-1262(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Reichwald K., Thiesen J., Wiehe T., Kioschis P., Straetling W.H.,
 RA Rosenthal A., Platzer M.;
 RT "Comparative analysis of the methyl Cpg binding protein 2 locus in man
 RT and mouse reveals new untranslated sequences.";
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: CHROMOSOMAL PROTEIN THAT BINDS TO METHYLATED DNA. IT CAN
 CC BIND SPECIFICALLY TO A SINGLE METHYL-CPG PAIR. IT IS NOT
 CC INFLUENCED BY SEQUENCES FLANKING THE METHYL-CPGS. MEDIATES
 CC TRANSCRIPTIONAL REPRESSION THROUGH INTERACTION WITH HISTONE
 CC DEACETYLASE AND THE COREPRESSOR SIN3A (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: NUCLEAR. COLOCALIZED WITH METHYL-CPG IN THE
 CC GENOME.
 CC -!- SIMILARITY: CONTAINS (R)GRP(K) MOTIFS AND SPK MOTIFS THAT HAVE
 CC BEEN FOUND TO BIND TO MINOR GROOVE OF AT-RICH DNA.
 CC -!- SIMILARITY: CONTAINS 1 METHYL-BINDING DOMAIN (MBD).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF072251; AAC68880.1; -;
 DR EMBL; AJ132922; CAB46495.1; -;
 DR EMBL; AF121351; AAF22116.1; -;
 DR EMBL; AF158181; AAF33024.1; -;

CC -!- SUBUNIT: INTERACTS WITH RAD51.
CC -!- TISSUE SPECIFICITY: HIGHEST LEVELS OF EXPRESSION IN BREAST AND
CC THYMUS, WITH SLIGHTLY LOWER LEVELS IN LUNG, OVARY, AND SPLEEN.
CC -!- DISEASE: BREAST CANCER (BC) IS AN EXTREMELY COMMON MALIGNANCY,
CC AFFECTING ONE IN EIGHT WOMEN DURING THEIR LIFETIME. A POSITIVE
CC FAMILY HISTORY HAS BEEN IDENTIFIED AS MAJOR CONTRIBUTOR TO RISK OF
CC DEVELOPMENT OF THE DISEASE, AND THIS LINK IS STRIKING FOR EARLY-
CC ONSET BREAST CANCER. MUTATIONS IN BRCA2 ARE THOUGHT TO BE
CC RESPONSIBLE FOR SOME INHERITED BREAST CANCER. IT IS LINKED WITH
CC MALE BREAST CANCER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

DR EMBL; X95152; CAA64484.1; .
DR EMBL; X95153; CAA64484.1; JOINED.
DR EMBL; X95154; CAA64484.1; JOINED.
DR EMBL; X95155; CAA64484.1; JOINED.
DR EMBL; X95156; CAA64484.1; JOINED.
DR EMBL; X95157; CAA64484.1; JOINED.
DR EMBL; X95158; CAA64484.1; JOINED.
DR EMBL; X95159; CAA64484.1; JOINED.
DR EMBL; X95160; CAA64484.1; JOINED.
DR EMBL; X95161; CAA64484.1; JOINED.
DR EMBL; X95162; CAA64484.1; JOINED.
DR EMBL; X95163; CAA64484.1; JOINED.
DR EMBL; X95164; CAA64484.1; JOINED.
DR EMBL; X95165; CAA64484.1; JOINED.
DR EMBL; X95166; CAA64484.1; JOINED.
DR EMBL; X95167; CAA64484.1; JOINED.
DR EMBL; X95168; CAA64484.1; JOINED.
DR EMBL; X95169; CAA64484.1; JOINED.
DR EMBL; X95170; CAA64484.1; JOINED.
DR EMBL; X95171; CAA64484.1; JOINED.
DR EMBL; X95172; CAA64484.1; JOINED.
DR EMBL; X95173; CAA64484.1; JOINED.
DR EMBL; X95174; CAA64484.1; JOINED.
DR EMBL; X95175; CAA64484.1; JOINED.
DR EMBL; X95176; CAA64484.1; JOINED.
DR EMBL; X95177; CAA64484.1; JOINED.
DR EMBL; U43746; AAB07223.1; .
DR EMBL; 274739; CAA98995.1; .
DR EMBL; 273359; CAA97728.1; .
DR MIM; 600185; .
DR InterPro; IPR002093; BRCA2_repeat.
DR Pfam; PF00634; BRCA2_repeat; 8.
DR PROSITE; PS50138; BRCA2_REPEAT; 8.

Query Match 4.7%; Score 145; DB 1; Length 3418;
Best Local Similarity 20.9%; Pred. No. 0.39;
Matches 96; Conservative 67; Mismatches 158; Indels 138; Gaps 23;
QY 36 KEDVAMELERVEDERPMWIKRSSECNPLQPIASAOFGATAGTCRKSYPGCGWVVK 95
DB 972 KSDISUNIDKIEKNDYMNKAG-----LGPISNHEGSGSFRITASNKEIKLSEHNKK 1026
QY 96 QRLFGKTAGRFVYFISPOGLKRSKSLIYLNKNGETSL-KPEDDFDTVLSSRGKISR 154
DB 1027 SKWFFK-----DIEEQYPTSL--ACVEIVNTLADNKKLSPQSIIN-TVSAHLQSSVV 1077
QY 155 YKDCSMAALTSL--QNSNNSNNWNLRTSKCKVDVFPMPSSSELQESRGLSNFTSTHL 212
DB 1078 VSDCKNSHITPQMLFSKODFNSNHL-----TPSOKAEITE-----LSTIL 1118
QY 213 LKKEDEGVDDVNFVRVK-----PKGKVTILK----- 239
DB 1119 -----ESGSQFEFTQFRKPSYILQKSTFEVPENQMTILKTSECRDADLHVIMNPSIG 1174

QY 240 -----GIPIKK-----TKKCKKSCSGFVQSDSKRESVCNKADAESEFPAOKSQL 284
DB 1175 QVSSKQFEGTVEIKRKFAGLLKNDCKNSASGYLTDE-----NEVGRGFYSAHGTKL 1227
QY 285 D-RTVCISDAGACGETLSVTSEENSLVKKKERSLSGSGNFCSEQKTSGLINKKFCSAKDS 343
DB 1228 NVSTEALQKAVKLFSDIENISEETS---AEVHPSLSSSKCHD---SVVSMF---KIEH 1277
QY 344 HNEKYEDTFLESEE-----IGPKVEVVERKEHLHPTDILKRGSEMNN--CSP 388
DB 1278 HNDK---TVSEKNKCOLLQNNIEMTTGTFFVEI-----TENYKRNTEEDNRYTAA 1327
QY 389 TRK----DFTGEKIFQEDTTPRTQIERKTSLYFSSKYN 423
DB 1328 SRSHNLEFGSGSSSKNDTV---CIHKDETDLFTDOHN 1363
RESULT 6
ID GAR2_SCHPO STANDARD; PKT; 500 AA.
AC P41831; O13707;
DT 01-NOV-1995 (Rel. 32, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN GAR2.
GN GAR2 OR SPAC140.02.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=95319932; PubMed=7596817;
RA Gulli M.-P., Girard J.-P., Zabetakis D., Lapeyre B., Melese T.,
RA Caizergues-Ferrir M.;
RT "gar2 is a nucleolar protein from Schizosaccharomyces pombe required
RT for 18S rRNA and 40S ribosomal subunit accumulation.";
RL Nucleic Acids Res. 23:1912-1918(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Brown D., Churcher C.M., Barrell B.G., Rajandream M.A., Wood V.;
RA Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RL -!- FUNCTION: HELPS THE ASSEMBLY OF PRE-RIBOSOMAL PARTICLES
CC CONTAINING 18S RNA.
CC -!- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.
CC -!- SIMILARITY: BELONGS TO THE GAR FAMILY.
CC -!- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; 248166; CAA88179.1; .
CC EMBL; AL163191; CAB86413.1; .
CC HSSP; P09651; 1HA1.
CC InterPro; IPR000504; RRM.
CC Pfam; PF00076; rrm; 2.
CC SMART; SM00360; RRM; 2.
CC PROSITE; PS0102; RRM; 2.
CC PROSITE; PS00030; RRM_RNP_1; 2.
CC Ribosome biogenesis; RNA-binding; Nuclear protein; Repeat;
CC rRNA processing.
CC FT DOMAIN 81 236 GLU/SER-RICH.
CC FT DOMAIN 263 341 RNA-BINDING (RRM) 1.
CC FT DOMAIN 366 443 RNA-BINDING (RRM) 2.
CC FT CONFLICT 339 339 S -> P (IN REF. 1).


```

CC CC -!- DEVELOPMENTAL STAGE: SPECIFIC TO LATE SPERMATOGENESIS.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC DR EMBL; Z22780; CAA80457.1; -
CC DR PIR; S35920; S35920.
CC DR PIR; B40713; B40713.
CC DR MIM; 603121; -
CC KW Cytoskeleton; Structural protein; Repeat; Sperm; Spermatogenesis.
CC FT NON_TER 1
CC FT DOMAIN 1 1 8 APPROXIMATE TANDEM REPEATS.
CC FT REPEAT 225 500
CC FT REPEAT 225 252
CC FT REPEAT 253 289
CC FT REPEAT 290 326
CC FT REPEAT 327 364
CC FT REPEAT 365 400
CC FT REPEAT 401 438
CC FT REPEAT 439 478
CC FT REPEAT 479 500
CC FT REPEAT 548 598
CC FT DOMAIN 1 1 PRO-RICH
CC SEQUENCE 598 AA; 68034 MW; 062BAZE2D2AB61F7 CRC64;
CC
CC Query Match 4.5%; Score 136.5; DB 1; Length 598;
CC Best Local Similarity 23.3%; Pred. No. 0.14;
CC Matches 111; Conservative 72; Mismatches 205; Indels 89; Gaps 21;
CC
CC QY 2 GTTGLLESISLGRGAAPTYSERLVPDPDLKEDVAMELERVEDEEQMMIKRSEC 61
CC Db 70 GGPPLKDKSKKGGSTATNPSKQVEETKTRQNEAD-KTPLKSSHENEQSKSSSET 128
CC
CC QY 62 NPLLQEPITASAGATAGTECKRSVPCGWVVRVVKORLFGKTAGRFVDVFI-----SPOGLK 117
CC Db 129 NP-----ESQSKTVSKNC50-----KDKSKSKSKKTNTFELTKNPKKDL 171
CC
CC QY 118 FRSKSLANLYLKHNGTSLKPEDFTVLKGRKISRYKDCSMAALTSHLON--QSNNSN 175
CC Db 172 KRKSTNDPISICSENSL---NVDFLVLGO-----SDDESINFDLAWLRYSNNKSN 222
CC
CC QY 176 WNLRTSKCKKDV--FMPPPSSSELQESR-GLSNFTSTHLLKKEDEGVDDV----- 223
CC Db 223 YSLKYTKYTKTKTKNAKSSDAESDSKADAKSKVKVKKDDKKDKKDKKDFESTDA 282
CC
CC QY 224 -----NFRKVRKPKGVITLKGIPKTKK-----GCRKSCSGFVQS 260
CC Db 283 ESGDSKDERKDTK-KDKKLLKDKKDKTKKYPSTDTESGDADKARNDSRLKASKND 341
CC
CC QY 261 DSKRES--VCNKADAEPVAKQSOLDRTVCISDAGAGETLSVTSENSLVKKKERSLS 318
CC Db 342 DKKDKAKITFTDSESELESSESKQDKDKDK-----SKTDNKKK-VKNDSESD 392
CC
CC QY 319 SGSNFCSEQKSGIITKFCSAKDSHNEKYEDT--FLESEIGTKVEVVERKEHLHTDIL 376
CC Db 393 ADSEPKGDSK-KGKDKERKGGKDKDKDKDAKNAESTEMESDLELKKDKKSKSE--- 448
CC
CC QY 377 KRGSEMDNCSPTKDFTEGKEIFQEDTTPRQIERKRTSLFYSSKYNEALSPPRK 433
CC Db 449 KRGSKDKIK-KDARKD-TESTDAEFDESSKTGF-KTSTKIKGSDTESSESLYKPGAK 502
CC
CC RESULT 9
CC BUD4_YEAST
CC ID BUD4_YEAST STANDARD; PRT; 1447 AA.
CC AC P47136;
CC DT 01-FEB-1996 (Rel. 33, Created)
CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
CC DT 30-MAY-2000 (Rel. 39, Last annotation update)

```

```

DE GN RUD SITE SELECTION PROTEIN BUD4.
OS BUD4 OR YJR052W OR J1905.
OC Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_taxid=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96295555; PubMed=8707826;
RA Sanders S.L., Herskowitz I.;
RT "The BUD4 prtein of yeast, required for axial budding, is localized
RL to the mother/BUD neck in a cell cycle-dependent manner.";
RN [2]
RP SEQUENCE FROM N.A.
RX Ramezani Rad M., Kirchrath L., Hollenberg C.P.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CO-ASSEMBLES WITH BUD3 AT BUD SITES. BUD4 AND BUD3 MAY
CC COOPERATE TO RECOGNIZE A SPATIAL LANDMARK (THE NECK FILAMENTS)
CC DURING MITOSIS AND THEY SUBSEQUENTLY BECOME A LANDMARK FOR
CC ESTABLISHING THE AXIAL BUDDING PATTERN IN G1.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN THE N-TERMINAL
CC SECTION DUE TO FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U41641; AAB17116.1; -
CC DR EMBL; Z49592; CAA89620.1; ALT_FRAME.
CC DR EMBL; Z49591; CAA89619.1; -
CC DR SGD; S0003852; BUD4.
CC DR InterPro; IPR001849; PH.
CC DR Pfam; PF00169; PH; 1.
CC DR SMART; SM00233; PH; 1.
CC DR PROSITE; PS50003; PH_DOMAIN; 1.
CC KW ATP-binding.
CC FT DOMAIN 181 189 ASP/GLU-RICH (ACIDIC).
CC FT NP_BIND 1175 1182 ATP (POTENTIAL).
CC FT DOMAIN 1302 1413 PH.
CC FT CONFLICT 291 328
CC FT CONFLICT 340 340
CC FT SEQUENCE 1447 AA; 164485 MW; 1D056B9FF1B7067B CRC64;
CC
CC Query Match 4.4%; Score 133; DB 1; Length 1447;
CC Best Local Similarity 19.5%; Pred. No. 0.73;
CC Matches 135; Conservative 105; Mismatches 269; Indels 184; Gaps 30;
CC
CC QY 1 MGTGLLESISLGRGAAPTYSERL-----VDPDPDLKEDVAM 41
CC Db 74 MSTISNESLMLGLLRNVSSELESPPAAVHOERTKNSVANGALGHANSPKVLNLLK--NMAQ 131
CC
CC QY 42 ELERVGEDEEQMMIKRKS-----ECNPLLQEPITASAGATAGTECKRSVPCGWVVRV 94
CC Db 132 DIDKLARDEKPKVLSLSSPLKFTLKSTQPLLSYPESPIH-----RSSIE-----I 176
CC
CC QY 95 KORLFKTAGRFVDVFI---SPOGLKFRKSSSLANLYLKHNG---ETSLKPEDFTVLKSK 148
CC Db 177 ETNYDDEDEEDDAYTCLTQSTQILHSPSRIPITNAVSINKLMLDFTLNPESDKSLVD 236
CC
CC QY 149 RGIKSKYKD-----CSMAA-----LTSHLONNSNNNRLTRSKCKK 186
CC Db 237 TSVDSGTGRELDITKTIPELPCKNSSTPEMTFVDEKCNLPKSLINTSNNSHDSRSPASVE 296
CC
CC QY 187 DV----FMPPPSSSELQESKGLSN--FTSHILLKED-----EGVDYVNRKVRKP 231

```

Db 297 DLNISTNLPAGDSSQ-----NNPVTDDADALITENDVVRDLQNMHEHIDDAFDEKVKLD 349
QY 232 KG-----KVTIL---KGIPKTKGCRKSCSGFVQOSKRES-----VCNKA 271
Db 350 ECCSNPEVTFLENDTRISIVYNSKGTANVQEFQEDSLAHSEPKFKDLNATSDVWVNE- 408
QY 272 DAESPVAQKSQLDRTVCISDAGACGETLSVTSENSLVKKERSLSGSGNFCSEOKTSG 331
Db 409 DKETDANISTSKSESYIAD-----YKTRQEDWDTK-----LHQSEHANEQ--PA 455
QY 332 TINKFCSAKSEHNEKYEDTLESEIGTKVVEVVERKEHLHTDLK-----RGSEM-- 382
Db 456 IIPQDSSEETTELNNSEFORNKDGEYRIVQHEESLYGORTKSPENLINGSEIGV 515
QY 383 -----DNCSPTKDFTEKIFQEDTTPRTQIERKTSL----- 416
Db 516 DHGEAAVNEPLAKTSABEHLSDSCEDQSVSEARNKDRIEKEVEVKDENITEKDESE 575
QY 417 YFSSKYNKEALSPRRKAKFKWTPPRSPFNVLQVETLFDHPW-----KLLIATIFLNRTSGK 472
Db 576 YHKVEENSEPEHVPLPLPW-----EELQNEPFIENDTNSIDLTRSKWP 625
QY 473 MAIPVLWKFLE---KYPSAEVARTADMRDVSSELLKPLGLYDLR--AKTIKVFSEYLYTKQ 527
Db 626 SDYISNIHQEEIKNSPESIANQSFSQSSITTASTVDSKKDNGSTSPFKPKPRIVRS 685
QY 528 WKY--PIELHGIGKYGNDSYRIFCVNEWKVQHP 558
Db 686 RIYNPKRSVSLNYDNEY-ILSNSEWNALDP 717

RESULT 10
BIRL_SCHPO STANDARD; PRG: 997 AA.
ID BIRL_SCHPO
AC 014064;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE BIRL PROTEIN (CHROMOSOME SEGREGATION PROTEIN CUT17).
GN BIRL OR CUT17 OR SPCC962.02C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID:4896;
RN [1]
RP SEQUENCE FROM N.A.
RA Morishita J., Matsusaka T., Yanagida M.
RT "Fission yeast cut17 is required for chromosome segregation.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Harris D., Wood V., Rajandream M.A., Barrell B.G.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=99398681; PubMed=10468581;
RA Uren A.G., Beilharz T., O'Connell M.J., Bugg S.J., van Driel R.,
RA Vaux D.L., Lithgow T.;
RT "Role for yeast inhibitor of apoptosis (IAP)-like proteins in cell
RT division";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10170-10175(1999).
CC -!- FUNCTION: SEEMS TO ACT IN THE PLEIOTROPIC CONTROL OF CELL
CC DIVISION. MAY PARTICIPATE IN CHROMOSOME SEGREGATION EVENTS.
CC -!- SIMILARITY: CONTAINS 2 BIR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AR031034: BAA83415.1; -;
DR EMBL: AL031323: CAA20434.1; -;
DR InterPro: IPR001370; BIR.
DR Pfam: PF00653; BIR: 2.
DR SMART: SM00238; BIR: 2.
DR PROSITE: PS01282; BIR_REPEAT_1; FALSE_NEG.
DR PROSITE: PS0143; BIR_REPEAT_2; 2.
KW Cell division; Repeat.
FT REPEAT 25 99 BIR 1.
FT REPEAT 120 194 BIR 2.
FT DOMAIN 80 83 POLY-ASP.
FT DOMAIN 312 319 POLY-ASP.
FT DOMAIN 487 490 POLY-SER.
SQ SEQUENCE 997 AA; 112579 MW; 952A6BAFA5C489F4 CRC64;

Query Match 4.3%; Score 131.5; DB 1; Length 997;
Best Local Similarity 20.1%; Pred. No. 0.56;
Matches 129; Conservative 119; Mismatches 241; Indels 153; Gaps 33;

QY 20 VTSSERLVPDPNDLRKEDVAMEL---ERVGEDEEQMMIKRS---SECNPLLQEPPIASA 72
Db 416 ISSSVSVGKEQNHTEKQ-VAIETPEQKVEKEDHLNLQGSFIEESTKQPISSKPTSS 474
QY 73 QGATAGTECRKSVPCGWERVVKQRLFGKTAGTDFYFISPGQLKPKRSKSSLIANYLHKG 132
Db 475 PDMTDAATGGRVSSSFRDKILOTFNFSRST-----IDFSNISKRNSEAN--DEND 526
QY 133 ETSLK---PE---DEFDTLSKRGIKSRVKDC-----SMAAL-----TSHLQNSNN 173
Db 527 ETNLKIPPEKRRKKEVQLSKNLVSSSTEDSHEPKVVTEDSQTALHVKFDELENKSM 586
QY 174 SNWNLTRSRCKKD---VMPSPSSSELSQES-----RGLSNFTSTHLLKEDGVDD 222
Db 587 SEQSLQSLSENDKPLDILPLAIAKRDNLVSGVLEKKGKSTSK---TKFDTSIVD 643
QY 223 VNFRRKPKGKVTILKGIPIKTKKGRKSCSGFVQ-----SDSKRESV 267
Db 644 F----IEKPKTEIS--EVLPEEKRAICDESQTVRVSIDRGVTKTROVSSPVSDKSENV 697
QY 268 CNKADAES-----EPVAKSOLDRTVCISDA--GACGETLSVTSENSLVKK 312
Db 698 -NHEEANGHTVMNVHSSLDPPQPIVQPNLESGSYLKDIPDRNVGNSEKVTFOEDDINS 756
QY 313 KERSLSSGNSFCSEQKTSIGIKFCSAKDEHNEKYEDTLESEIGTKVVEVVERKEHLH 372
Db 757 KLSKNQTVAEAVNTETSDKLOKEANHELENIEKIEEKLETVDKVLSLSDAPDQEIKN 816
QY 373 TDILKRGSEMDNNSPTKDFTEKIFQEDTTPRTQIERR-----KTSLYFSSKY-NKEAL 427
Db 817 RTSVQNGTRSVSKNTPEKE---TKVDKIDNVSKKDVETSPGSCETSSAFKTYAEKVT 872
QY 428 S----PPRRKAFKK--WTPPRSPFNVLQVETLFDHPWKLILLATIFLNKTSKMAIPVLWKL 482
Db 873 SINLPSVRKPLDESYDHSISPFDPCLQSSFLAP-----QTPVKSCKHALPLV--- 919
QY 483 EKYPSEAEVARTADWR--DVSELLK-----PLGLYDLRAKTIVKFSDEYLTQW-KYPIEL 534
Db 920 -----EANAPWPEIDFSSLLSESPVPNPVFNPKLSEKEL-----DMTVEQWIKF--- 963
QY 535 HGIGKYGNDSYRIFCVNEWKVQHPEDHKLNKYHDLWLENHEK 576
Db 964 ---MVAK-----CAKEFEACEE--KI---EWLLECKR 989

RESULT 11
ID ANK2_HUMAN STANDARD; PRG: 3924 AA.
AC Q01484; Q01485;
DT 01-APR-1993 (Rel. 25, Created)

Db 9 VTPSGKLVSETVKNKETSDEETLK-----AKGKIKVNSRREDFKQKQP 59
 QY 165 SH-----LQNSNSNNLRTSKCKKDVMPSPSS-----SELOE----- 200
 Db 60 SKKKRIIYDSSEETLOVKNNAKPPKLPVSSRPGKLSRDPVTYISETDEEDFMCK 119
 QY 201 -----SRGLSNFTSHL-----LLKDEGVDDVNFVKRKPCKVILKGIPIKTKKCR 251
 Db 120 KAASKKNGRSTNSHLGTSNKKKEE-----NTRTKNKPUS-----PIKLTPTSVL 166
 QY 252 KSC-SGQVQSD-----SKREVCNKA-----AESEPPVAQKSQSDRTV-----CISDAGACG 297
 Db 167 DVFGTGVSQVRSNKKMVASRKRKLSQNTDSGLNDEAIAKQLQDDEAELEQLHEDEFA 226
 QY 298 ETLVSTSENSLVKKERSLSSGNSFCSEQ-----KTSGLI----- 333
 Db 227 RTLAMDDEPK-TTKARKDTEAGETFSVQANLSKAEKHYPHKYKTAQVSDERKSYSPR 285
 QY 334 --NKFCASDSE-EHNEKYEDTF--LESEIGTKVFEVVERKEH-LHTDILKRGSEMDNNC- 386
 Db 286 KQSKYESSKESQOHSKSSADKIGEVSSPKASSKLAIMKRKESSYKEIEPVASKRKENAI 345
 QY 387 -----SPTKDFTEKIFQEDTTPRTQIERKTSLY-FSKYNKEALSPPR 431
 Db 346 KLGKETKTPKTKSSPAK-----ESVSPDS-----EKKRTNYQARYSLNREG----- 390
 QY 432 RKAFKWTTPRSPFNLVQ-----ETLPHDPWKLIIAT-----IFL 466
 Db 391 PKALGSKEIPKGAENCLGELIFVITCVLESIERDEAKSLIERGGKVTGNYSKKTNYLVN 450
 QY 467 NRTSGK-----MAIPVLWKELE-----KYPSEAVARTADWDVSEL----- 502
 Db 451 GRDSGQSKDKAAALGTKIIDEGLNLRTMPGKKSKYEIAVETEMKESKERTPOKN 510
 QY 503 -----LKPGLYDLRAKTIIVKPSDEVLTKQWK----- 529
 Db 511 VQGRKISPSKESSEKSRPTSKRDSLAKTIKETDVF-----WKSDFKEQVAETSGD 566
 QY 530 -----YPIELHG-IGKYGNDSDYRIFVNE-----WKQVH 557
 Db 567 SKARNLADDSSENKVENLLWVDKYKPTSLKTIIGOGDQS-----CANKILLMLRNWQKSS 622
 QY 558 PEDHKLNY 566
 Db 623 SEDKHKSKF 631

RESULT 13

YF16_YEAST STANDARD; PRT; 1233 AA.
 AC F43597;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE HYPOTHETICAL 137.7 KDA PROTEIN IN UGSI-FAB1 INTERGENIC REGION.
 GN YF016C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 ON NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RX MEDLINE=95400292; PubMed=7670463;
 RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
 RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
 RA Yamazaki M., Tashiro H., Eki T.;
 RT "Analysis of the nucleotide sequence of chromosome VI from
 RT Saccharomyces cerevisiae."
 RL Nat. Genet. 10:261-268(1995).
 CC -!- SIMILARITY: SOME, TO MAMMALIAN NEUROFILAMENT TRIPLET M PROTEIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D50617; BAA09255.1; -
 DR SGBL: S0001912; YF016C.
 KW Hypothetical protein.
 SQ SEQUENCE 1233 AA; 137697 MW; C8A7CD2C6F0892F6 CRC64;

Query Match 3.9%; Score 120.5; DB 1; Length 1233;

Best Local Similarity 18.9%; Pred. No. 3.7;
 Matches 121; Conservative 91; Mismatches 228; Indels 199; Gaps 25;

QY 7 ELSLGRGAAPVITS-----SERLVDPDPND-----LRKEDVAMELERYGEDEQM 53
 Db 546 DNLVLEDEAEAPTQENKPTVEVGEIDIPADPRDDVLEAVEKNIIPEDLEVAKEDEGE 605
 QY 54 MIKRSSECNPLQPEPI-----ASAQFGATA- 78
 Db 606 QVK-----LDPEVAMKDDKIAMRGAESISSEDMKKKQEGTAELSNEKAKKEDVETAR 657
 QY 79 ---GTECRKSVPCGWERYVQKRLFGKTAGRDYVYFISPOGLKPRS----- 120
 Db 658 ESAGVEVEKSKTPSPKVVKR-----CTSGK-----PEDLQINERDPEVLKEDVRVPD 706
 QY 121 ---KSSILANYLHKNE-----TSL-----KPEDFDTV 145
 Db 707 EDVKEPATIENSEEEDPKSQVQISTEQAEITQKMDGVGTTSFKEEEKPKRFEIT- 765
 QY 146 LSKRGIKSYKDCSMAALTSHLQNSNN-SNNLNLRTSKCKKDVMPSPSSSELOESRGL 204
 Db 766 --QEGDKITCKD-----TNHEGCEATEAASSENSKASDVGTAEKYIDPSSSVKDDT--- 814
 QY 205 SNFTSTHLLKDEGVDDVNFVKRKPCKVITLKGPIKTKKCKKSCGFVQSDSKR 264
 Db 815 -----EEDAEVENSEKTEFIKVAE-----LENLDAPKAE-----VTAEINK 852
 QY 265 ESVCKADAESEPAQKSQSDRTVCIS-----DAGAGETLSVTSSENSLVKKKRSLS 318
 Db 853 ENEDVEVDTEDAEVENSEKTEFIKVAELNLDAPKAEVTAELNKENEDVEVAATSKE 912
 QY 319 SGSNFCSEQTSIGTIKFC-----SAKDSHENEKYEDTFLESEIGTKVFEVVERKEHLH- 372
 Db 913 DIETKCEPAETPIEDGTCTEAEVSKKDAEAVTK-EDENMENSKEIAEALKDVTGDQDIDD 971
 QY 373 --TDILKRGSEMDNCSPTKDFTG-----EKIFQEDTIPRTQIERKTSLYFSSKY 422
 Db 972 INISDEFQRTVELPELEKQDIKKNKGEKDELEVEETKETSPLDVLVEENITEKNEIKQ 1031
 QY 423 NKEALSP---PRRKAPKWTTPRSPFNLVQETLPHDPWKLIIATIF---LNRTSGKMAIP 476
 Db 1032 EEEVSQLDFNETESISKEAPNNDENFEGDQSTRENPKASADDPKIDILDETN----- 1085
 QY 477 VLVKELEYPSAEVARTADWDVNSSELLKPLGLYDLRAKT 515
 Db 1086 ---EFLEQ-----LKIVDDSENLALQSLDAKDDSTTQT 1115

RESULT 14

AC15_MOUSE
 ID AC15_MOUSE STANDARD; PRT; 1131 AA.
 AC P35601;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ACTIVATOR 1 140 KDA SUBUNIT (REPLICATION FACTOR C LARGE SUBUNIT) (A1
 DE 140 KDA SUBUNIT) (RF-C 140 KDA SUBUNIT) (ACTIVATOR 1 LARGE SUBUNIT)
 DE (A1-P145) (DIFFERENTIATION SPECIFIC ELEMENT BINDING PROTEIN)

DE (ISRE-BINDING PROTEIN).
GN RFCL OR RECC1 OR IBF-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BA1B/C;
RX MEDLINE=94089669; PubMed=8265586;
RA Burdello P.D., Utani A., Pan Z., Yamada Y.;
RT "Cloning of the large subunit of activator 1 (replication factor C)
RT reveals homology with bacterial DNA ligases.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:11543-11547(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94158835; PubMed=8114700;
RA Luckow B., Bunz F., Stillman B., Lichter P., Schuetz G.;
RT "Cloning, expression, and chromosomal localization of the
RT 140-kilodalton subunit of replication factor C from mice and
RT humans.";
RL Mol. Cell. Biol. 14:1636-1634(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS;
RX MEDLINE=95388065; PubMed=7659092;
RA McGehee Habener J.F.;
RT "Differentiation-specific element binding protein (DSEB) binds to a
RT defined element in the promoter of the angiotensinogen gene required
RT for the irreversible induction of gene expression during
RT differentiation of 3T3-L1 adipoblasts to adipocytes.";
RL Mol. Endocrinol. 9:487-501(1995).
RN [4]
RP SEQUENCE FROM N.A.
RA Haque S.J.;
RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-565 FROM N.A.
RC STRAIN=LAF1;
RA Lossie A.C., Haugen B.H., Wood W.M., Camper S.A., Gordon D.F.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 354-528 FROM N.A.
RX MEDLINE=90229765; PubMed=1691767;
RA Haque S.J., Kumar A., Fischer T., Rutherford M.N., Williams B.R.;
RT "Evaluation of inter- and intramolecular primary structure homologies
RT of interferons by a Monte Carlo method.";
RL J. Interferon Res. 10:31-31(1990).
CC -!- FUNCTION: THE ELONGATION OF PRIMED DNA TEMPLATES BY DNA POLYMERASE
CC DELTA AND EPSILON REQUIRES THE ACTION OF THE ACCESSORY PROTEINS
CC PCNA AND ACTIVATOR 1. THE 140 SUBUNIT BINDS TO THE PRIMER-
CC TEMPLATE JUNCTION.
CC -!- SUBUNIT: HETEROPENTAMER OF SUBUNITS OF 140/145, 40, 38, 37, AND
CC 36.5 KDA THAT FORMS A COMPLEX WITH PCNA IN THE PRESENCE OF ATP.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: BELONGS TO THE ACTIVATOR 1 140 KDA SUBUNIT FAMILY.
CC -!- SIMILARITY: CONTAINS 1 BCT DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U01222; AAA21643.1;
CC EMBL: X72711; CAA51260.1;
CC EMBL: U36441; AAA79698.1;
CC EMBL: U07157; AAC52140.1;
CC EMBL: U15037; AAB60452.1;
CC MGD; MGI:97891; Reccl.

DR InterPro: IPR001939; AAA_subfam.
DR InterPro: IPR001357; BRCT.
DR InterPro: IPR000862; RFC.
DR Pfam: PF00004; AAA; 1.
DR Pfam: PF00533; BRCT; 1.
DR SMART: SM00292; BRCT; 1.
DR PROSITE: PS0172; BRCT; 1.
KW DNA replication; ATP-binding; Transcription regulation; DNA-binding;
KW Activator; Nuclear protein.
FT DOMAIN 399 477 BRCT.
FT NP_BIND 635 642
FT ZN_FING 734 751
FT DOMAIN 1104 1108
FT DOMAIN 354 528
FT
FT CONFLICT 66 66
FT CONFLICT 187 187
FT CONFLICT 254 254
FT CONFLICT 559 559
FT CONFLICT 614 614
FT CONFLICT 945 945
FT CONFLICT 1071 1071
FT CONFLICT 1104 1104
FT CONFLICT 1131 1131
SQ SEQUENCE 1131 AA; 125984 MW; A6F4F970A7F9EE94 CRC64;

Query Match 3.9%; Score 120; DB 1; Length 1131;
Best local Similarity 20.1%; Pred. No. 3.6;
Matches 120; Conservative 76; Mismatches 182; Indels 218; Gaps 30;
QY 119 RSKSSLLYLNKNGETSLK-----PEDFDFTLSKRGKISRYKDCSMAALTSHLQN 169
DB 82 KKKSEKLSLYKPGKVSQKDPVTVSETDEDDF-VCKKAASKSKENGVS-----TNSYLG 136
QY 170 QSN--NSNWNLTRSKCKKDFMFPSS-----SSSELOQSR 202
DB 137 TSNVKKNEENVKTKNKLPTSTVLDVFGTESVORSGKKMVTSKRKESQNTEDSR 196
QY 203 GLSNFTSHILLKKEGV-----DDVNFKV-----RKPQKVTILKIPKTKKCRK 252
DB 197 LNDEAIAKQQLDDEDAELERQLHIEDEFARTLALLDEPK-----IKARKDSEE 246
QY 253 SCSGF--VQSD---SKRESVCNKAD-----AESEPVAV--KSLQDRTV 288
DB 247 GEESFSSVQDDLSKAKOKSPNKAELFSTARKTYSAPAKHGKGRASEDAKOPCKSAHRKEA 306
QY 289 CTSAGACGETLSVTSENS-----LVKKKRSLSGNSFCSEQKT--SGIINKFCSA 339
DB 307 CSSPKASAKLALMAKAESESYNETELLAARRKESATEPKGKTKTPKTKVSPTKRESVSP 366
QY 340 KDSH---NEKYDTFLESE---EIGTK-----VEVVERKEHLHT 373
DB 367 EDSKKRTNYQARYSLNREGPKALGSKETPKGAENCLGLTFVTGVLESIERDE-AKS 425
QY 374 DILKRGSEMDNCS-----PTRKD---FTGEKIFQD-----TIP----- 405
DB 426 LIERYGKVTGVNSKNTNYLVMGDRSQSKSDKAAALGTLKILDEGLDLIRTPGKRK 485
QY 406 -----RTQIERKTSYFSSKYN---KEALSPRR-----RAFKKWTPPRPFNLVQETLF 453
DB 486 YEMAAEAEKMKKSKLRTPOKNDQGRKISPARKKESKCKLTLKLNKSPMKAVKKEAS 545
QY 454 HDPKLLIATIFLRTSGKMAIPVLWKFLEKYPSEAVARTADWDVSELLKPLGLYDLRA 513
DB 546 TCPRLGVDKETHGRSKNKEEC-LLW--VDKY-----KPAKLKNI-- 582
QY 514 KTVKFSDEYLTQWKYPTELHGIGKYGNDYSRIFCVNE-----WKQVHPDHK 562
DB 583 -----IGQQDQS-----CANKLLRLNRLNWKSPPEEK 611

RESULT 15
YM67_YEAST

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 14:06:25 ; Search time 134.53 seconds
(without alignments)
344.669 Million cell updates/sec

Title: US-09-724-296-37

Perfect score: 1662

Sequence: 1 MIFRFGFVSNAMSLWDASPA.....ELSSIRGVKRIIGGALQWKS 317

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1001	60.2	317	2 Q914F0	Q914f0 bacillus ce
2	1001	60.2	325	2 Q914E9	Q914e9 bacillus ce
3	484.5	29.2	322	2 Q9K9P8	Q9K9p8 bacillus ha
4	359.5	21.6	599	3 Q10988	Q10988 schizosacch
5	345	20.8	656	3 Q01408	Q01408 neurospora
6	293.5	17.7	305	2 Q9S0M9	Q9S0m9 deinococcus
7	293.5	17.7	326	2 Q9RTE6	Q9rte6 deinococcus
8	109.5	6.6	695	3 Q9HGM8	Q9hgm8 schizosacch
9	104.5	6.3	315	2 Q9X2Y9	Q9x2y9 bacillus an
10	104.5	6.3	1690	3 Q9P977	Q9p977 candida alb
11	104	6.3	2272	5 Q9VZS7	Q9vzs7 drosophila
12	101.5	6.1	1690	3 Q9P411	Q9p411 candida alb
13	101.5	6.1	1690	3 Q9HFT8	Q9hft8 candida alb
14	101	283	1	P95930	P95930 sulfolobus
15	100	6.0	823	2 Q9KB81	Q9kb81 bacillus ha
16	98	5.9	331	2 Q4S152	Q4s152 bacteroides
17	98	5.9	505	5 Q9NGC2	Q9ngc2 tribolium c
18	97.5	5.9	1744	11 Q9R095	Q9r095 rattus norv
19	96.5	5.8	969	3 Q14014	Q14014 schizosacch

20	96.5	5.8	1760	5 Q9VLT3	Q9vlt3 drosophila
21	95	5.7	775	5 Q45519	Q45519 caenorhabdi
22	94.5	5.7	553	5 Q19702	Q19702 caenorhabdi
23	94.5	5.7	1145	10 Q04146	Q04146 acetabulari
24	94	5.7	553	2 Q9FCX5	Q9fcx5 clostridium
25	93.5	5.6	805	5 Q17235	Q17235 caenorhabdi
26	93.5	5.6	1066	5 Q9TZX4	Q9tzx4 caenorhabdi
27	93	5.6	674	5 Q20833	Q20833 caenorhabdi
28	92.5	5.6	234	4 Q9Y2M6	Q9y2m6 homo sapien
29	92.5	5.6	601	4 Q75226	Q75226 homo sapien
30	92.5	5.6	674	4 Q9UBU7	Q9ubu7 homo sapien
31	92	5.5	261	12 Q9YFX9	Q9ypx9 maize strea
32	92	5.5	428	1 Q9UYK0	Q9uyk0 pyrococcus
33	92	5.5	921	2 Q91Z27	Q91zz7 heisseria m
34	91.5	5.5	237	2 Q9PCN0	Q9pcn0 xyliella fas
35	91.5	5.5	805	5 Q93719	Q93719 caenorhabdi
36	91.5	5.5	1035	4 Q9H2F7	Q9h2f7 homo sapien
37	91.5	5.5	1052	11 Q9ERA1	Q9eral microtus ar
38	91.5	5.5	1409	5 Q08461	Q08461 drosophila
39	91	5.5	558	10 Q9STH1	Q9sthl arabidopsis
40	91	5.5	852	2 Q9K940	Q9k940 bacillus ha
41	91	5.5	1020	10 Q9ZVT9	Q9zvt9 arabidopsis
42	90.5	5.4	279	5 Q17470	Q17470 caenorhabdi
43	90.5	5.4	448	5 Q18125	Q18125 caenorhabdi
44	90.5	5.4	553	5 Q19881	Q19881 caenorhabdi
45	90.5	5.4	625	5 Q44733	Q44733 caenorhabdi

ALIGNMENTS

RESULT	1				
Q914F0					
ID	Q914F0	PRELIMINARY;	PRT;	317 AA.	
AC	Q914F0;				
DT	01-OCT-2000 (Tremblrel. 15, Created)				
DT	01-OCT-2000 (Tremblrel. 15, Last sequence update)				
DT	01-MAR-2001 (Tremblrel. 16, Last annotation update)				
DE	PUTATIVE UV-ENDORNUCLEASE.				
GN	YWJD.				
OS	Bacillus cereus.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;				
OX	Bacillus/Staphylococcus group; Bacillus.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-ATCC 14579 TYPE STRAIN;				
RX	MEDLIN-20055637; PubMed=10589720;				
RA	Okstad O., Gominet M., Purnelle B., Rose M., Lereclus D., Kolsto A.B.;				
RT	"Sequence analysis of three Bacillus cereus loci under PlcR-regulated				
RL	genes encoding degradative enzymes and enterotoxin."				
RL	Microbiology 145:3129-3138(1999).				
DR	EMBL; AJ243712; CAB69813.1;				
KW	Endonuclease.				
SQ	SEQUENCE 317 AA; 36990 MW; A6AC35F5800A22C3 CRC64;				

Query Match 60.2%; Score 1001; DB 2; Length 317;
Best Local Similarity 58.2%; Pred. No. 2.7e-74;
Matches 185; Conservative 55; Mismatches 74; Indels 4; Gaps 2;

OY	1	MIFRFGFVSNAMSLWDASPAKTLTFARYSKLSKTERKALLTVTKANLRNTRMTLHYIIG	60
DB	1	MLIRGGYSHAMALWDCSPAKTMTFTSPKLSKOEKDKLVHVRQNLERTIRILYNIA	60
OY	61	HGIPLYRESSSIVPLATHPDVMDVFTPFQKEFREIGELVKTHOLRTSFHPNQFTLTSP	120
DB	61	HEIPLYRLSSSIVPLATHPEVEFOYIGVFTPLWRKIGALIKEHNLRISFHPNQFTLTSD	120
OY	121	KESVTKNAVTOYMYHYRMLEAMGADRSVINIHGGVGNKDTATQAFQNIKOLPOEIK	180
DB	121	KPHITTNALTMDTYHYKILDAIGTADSSVINIHVGAGYGNKEKATERPHENIKKLPHIK	180

```

QY 181 ERMTELENDKTYTTEETLQVCEQEDVPVDFHHFYANPDHDLNVALPRMIKTWERIG 240
Db 181 KOMTELENDKTYTTEETLQVCEQEDVPVDFHHFYANPDHDLNVALPRMIKTWERIG 239
QY 241 LOPKVHLSPPKSEQAIRSHADYVDANFL---LERFROWGNTNIDPMIEAKOKDALLRLMD 297
Db 240 ISPKVHSSPSEKFEAFAEYIDLEFIKPLHTAKKHHNFIDIMIESKQKDLALFOLID 299
QY 298 ELSSIRGVKRIKGGALOW 315
Db 300 ELSAIRGIRKISGAMLOW 317

RESULT 2
Q9L4E9 PRELIMINARY; PRT; 325 AA.
AC Q9L4E9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE PUTATIVE UV-ENDORNUCLEASE.
GN YWJD.
OS Bacillus cereus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1396;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14579 TYPE STRAIN;
RX MEDLINE=20055637; PubMed=10589720;
RT "Sequence analysis of three Bacillus cereus loci under P1cr-regulated
RT genes encoding degradative enzymes and enterotoxin."
RL Microbiology 145:3129-3138(1999).
DR EMBL: AJ243712; CAB69814.1; -.
KW Endonuclease.
SQ SEQUENCE 325 AA; 37900 MW; 07FBDD1FA580653 CRC64;

Query Match 60.2%; Score 1001; DB 2; Length 325;
Best Local Similarity 58.2%; Pred. No. 2.8e-74;
Matches 185; Conservative 55; Mismatches 74; Indels 4; Gaps 2;

QY 1 MIFRPGFVSNAWLDASPAKTLTFARYSKLSKTEREALLTVTKANLRNTMRTLHIIG 60
Db 9 MLIRGYVSHAMALWDCSPAKTMTFTSFKLSKQERDKYHVIHQLEHTIRTLHYNIA 68
QY 61 HGIPLYRESSIVPLATHPDMVDFVTPFOKEFREIGELVKTHOLTSFHPNQFTLFTSP 120
Db 69 HEIPLYLSSIVPLATHPEVEFDYIGVFTPLWRKIGALIKEHLNLSFHPNQFTLFTSD 128
QY 121 KESVTKNAVTDMAHYRMLEANGIADRSVINIHGGAYGNKDTATATAPHONIKOLPOEIK 180
Db 129 KPHITNAVTDMAHYRMLEANGIADRSVINIHGGAYGNKDTATATAPHONIKOLPOEIK 188
QY 181 ERMTELENDKTYTTEETLQVCEQEDVPVDFHHFYANPDHDLNVALPRMIKTWERIG 240
Db 189 KOMTELENDKTYTTEETLQVCEQEDVPVDFHHFYANPDHDLNVALPRMIKTWERIG 247
QY 241 LOPKVHLSPPKSEQAIRSHADYVDANFL---LERFROWGNTNIDPMIEAKOKDALLRLMD 297
Db 248 ISPKVHSSPSEKFEAFAEYIDLEFIKPLHTAKKHHNFIDIMIESKQKDLALFOLID 307
QY 298 ELSSIRGVKRIKGGALOW 315
Db 308 ELSAIRGIRKISGAMLOW 325

RESULT 3
Q9K9P8 PRELIMINARY; PRT; 322 AA.
AC Q9K9P8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)

```

```

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE BH2597 PROTEIN.
GN BH2597.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL: AP001516; BAB06316.1; -.
KW Complete proteome.
SQ SEQUENCE 322 AA; 36900 MW; 4B0DC9291CD1E7B4 CRC64;

Query Match 29.2%; Score 484.5; DB 2; Length 322;
Best Local Similarity 32.2%; Pred. No. 6.7e-32;
Matches 106; Conservative 67; Mismatches 133; Indels 23; Gaps 6;

QY 1 MIFRPGFVSNAWLDASPAKTLTFARYSKLSKTE---RKEALLTVTKANLRNTMRTLHI 57
Db 1 MRIOQIVAMSELANASPKTMTATQFEKIEDHAGLRK--LERIAKTNLNLCLRLKH 58
QY 58 IIGHGIPLYRESSIVPLATHP-DVMWDFVTPFOKEFREIGELVKTHOLTSFHPNQFTL 116
Db 59 NLAYQISFTRLSKSLVPLVNHPLTEGKYELALAEALQAVGEFASEHOMRIDFHPDFV 118
QY 117 FTSPKESVTKNVTDMAHYRMLEANGIADRSVINIHGGAYGNKDTATATAPHONIKOLP 176
Db 119 LNSEAKEITRRSLQTLTYHYKLLKGMEIDPRHRCVLHVGGKKGVEAGLEQFIENTASIP 178
QY 177 QETKERMTLENDKTYTTEETLQVCEQEDVPVDFHHFYANPDH---ADLNVALP 230
Db 179 KSLLSMIMLENDKSYTIDDLVYLGEKLAIPVVDIHH-----HDVLHRSKSLQETWQ 231
QY 231 RMKTWERIGLOPKVHLSPPKSEQAIRSHADYVDANFLERFROWGT---NIDPMIEAK 286
Db 232 RIVATWEDSPLPKVHLSPLSGEDDPRHHYINADRFATLHEICADAVDHLHYMIEAK 291
QY 287 QKDKALLRLMDELSSIRGVKRIKGGALOW 315
Db 292 KKDLALFQLMKDLAEYDEITVVSXSAVEF 320

RESULT 4
Q10988 PRELIMINARY; PRT; 599 AA.
AC Q10988; P87339;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE UV-ENDORNUCLEASE.
GN UVDE OR UVEI+.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sp972;
RX MEDLINE=9618860; PubMed=8614629;
RA Takao M., Yonemasa R., Yamamoto K., Yasui A.;
RT "Characterization of a uv endonuclease gene from the fission yeast
RT Schizosaccharomyces pombe and its bacterial homolog."

```

Nucleic Acids Res. 24:1267-1271(1996).

RP SEQUENCE FROM N.A.
 RX MEDLINE=97175806; PubMed=9023111;
 RA Davey S., Nass M.L., Ferrer J.V., Sidik K., Eisenberger A.,
 RA Mitchell D.L., Freyer G.A.;
 RT "The fission yeast UVDR DNA repair pathway is inducible";
 RN Nucleic Acids Res. 25:1002-1008(1997).
 RP SEQUENCE FROM N.A.
 RC STRAIN=972H-;
 RA Lyne M., Rajandream M.A., Bartell B.G., Lucas M., Gaillardin C.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; D78571; BA011415.1;
 DR EMBL; U78487; AAC49664.1;
 DR EMBL; AL023859; CAA19577.1;
 KW Endonuclease.
 SQ SEQUENCE 599 AA; 68816 MW; C09ED7219D5895BE CRC64;

Query Match 21.6%; Score 359.5; DB 3; Length 599;
 Best Local Similarity 29.7%; Pred. No. 2.8e-21;
 Matches 83; Conservative 60; Mismatches 111; Indels 25; Gaps 7;

QY 34 TERKEALLTVTKANLRNTW---RTLHYLIIGHGIPLYRFSSTIVPLATHPDVMMDFVTFP- 89
 DB 276 TQROGLESVKQIGTONVLDLIKLEWNNFNGIHFMVRSSDLFPFASH--AKYGYTLEFA 333

QY 90 KBREFEIGELVTHQURTSFHPNQFTLTSPKESVTKNVDMAYHYRMLEAMGIADR-- 147
 DB 334 QSHLEEVGLKANKYHRLTMHPQVQTOIASPREVVVDSAIRDLAYHDEILSRMKLNEQLN 393

QY 148 --SVINIHGGAYGNKDTATAPHQNKOLPOEIKERMTLENDDKTYTTEETLVCEQED 205
 DB 394 KDAVLIHLGGTPEGRKETIDRFKRYORLSVSKARLVLENDVSVQDILLPLCQELN 453

QY 206 VPVFDFPHFYANP---DDHADNLVALPRMTKTWERIGLQPKVHLSSPKSEQAI----- 256
 DB 454 IPLVDMHHNNIVPGRLRGSLDMLPIITRETWTWKGITOKOHYSESADTAISGMKR 513

QY 257 RSHADYVDANFLLERFRQWGTNIDFMIEAKOKDKALLRL 295
 DB 514 RAHSDRV-----FDFPPCDPTMDLMEAKEKEQAVFEL 546

RESULT 5
 Q01408 PRELIMINARY; PRT; 656 AA.
 ID Q01408
 AC Q01408
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
 DE UV-ENONUCLEASE.
 GN UVEL.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=74-OR23-1A;
 RX MEDLINE=9522980; PubMed=7774597;
 RA Yajima H., Takao M., Yasuhira S., Zhao J.H., Ishii C., Inoue H.,
 RA Yasui A.;
 RT "A eukaryotic gene encoding an endonuclease that specifically repairs
 RT DNA damaged by ultraviolet light."
 RL EMBO J. 14:2393-2399(1995).
 DR EMBL; D11392; CAB20113.1;
 KW Endonuclease.
 SQ SEQUENCE 656 AA; 74413 MW; 4B231E7A8735C86F CRC64;

Query Match 20.8%; Score 345; DB 3; Length 656;

Best Local Similarity 32.3%; Pred. No. 4.9e-20;
 Matches 87; Conservative 51; Mismatches 105; Indels 26; Gaps 8;

QY 46 ANLNTMRTLHYIIGHGIPLYRFSSTIVPLATHPDVMMDFVTFPOKE-FREIGELVKTHQ 104
 DB 277 ANARDIVKMLCWNKBYGIRFLRSLSEMFPEASHP-VHGYKLAPFASVLAEGRAVAELG 335

QY 105 LRTSFHPNQFTLTSPKESVTKNVDMAYHYRMLEAMGIADR-----SVINIHGGAYGN 160
 DB 336 HRLTHPGQFTQLGSPKRVESAIRDLEHYDELILLSLLKLPQONRDVAVMIHMGQFGD 395

QY 161 KDTATAPHQNKOLPOEIKERMTLENDDKTYTTEETLVCEQEDVFPVDFHFIHYANPD 220
 DB 396 KAATLERFKRYARLSQCKNRLVLENDVGVTVHDLPLVCEELNIPMWLDYHHNHCIFD 455

QY 221 D-----HADNLVALP---RMKTWERIGLQPKVHLSSPKSEQAI-----BSHADYVDANF 267
 DB 456 PAHLREGTLDISDPKQERIANTWKRGIKOKMHYSEP-CDGAVTPRDKRRHRPRV----- 510

QY 268 LLERFRQWGTNIDFMIEAKOKDKALLRLM 296
 DB 511 --MTLPPCPDMDLMEAKEKEQAVFELM 537

RESULT 6
 Q09509 PRELIMINARY; PRT; 305 AA.
 ID Q09509
 AC Q09509
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
 DE UV-ENONUCLEASE.
 GN UVSCDE.
 OS Deinococcus radiodurans.
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KRI;
 RA Kitayama S., Kikuchi M., Funayama T., Narumi I., Watanabe H.;
 RT "Cloning of structural gene of an alternative inclusion enzyme for DNA
 RT damage in Deinococcus radiodurans."
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB033747; BAA85759.1;
 KW Endonuclease.
 SQ SEQUENCE 305 AA; 33592 MW; B94D333243E2FEA4 CRC64;

Query Match 17.7%; Score 293.5; DB 2; Length 305;
 Best Local Similarity 30.1%; Pred. No. 2.9e-16;
 Matches 88; Conservative 51; Mismatches 120; Indels 33; Gaps 11;

QY 21 KTLTFARYSKLSKTERKALLTVTKANLRNTMRTLHYIIGHGIPLYRFSSTIVPLATHPD 80
 DB 25 RTVTLRSYRALSPAREAKLLDLYSSNKTIRGAADYCAAHDIRLYLSSSLFPMPL---D 81

QY 81 VMWD-----FVTPFOKEFREIGELVKTHTLTSFHPNQFTLTSPKESVTKNVDMAYH 135
 DB 82 LAGDDTGAAVLTALPQLLEAGHAFTDAGVRLLMHPEQFIVLNSDRPEVRESSVRAMSAH 141

QY 136 YRMLEAMGIADRSVINIHGGAYGNKDTATAPHQNKOLPOEIKERMTLENDDKTYTTE 195
 DB 142 ARVMDGLGLA-RTPWNLL--LHGKGGKGAELAAIPDLDPVRLRLGLENDERAYS 198

QY 196 ETLVCEQEDVFPVDFHFIHYANP--DDHADNLVA--LPRMKTWERIGLQPK-----VHL 247
 DB 199 ELLPICEATGTPLVFDAAHHVVDKLPQEDPSVREWLRLARATW-----QPPEQVWHL 253

QY 248 SSPKSEQAIRSHADYVDANFLLERFRQWGTNIDFMIEAKOKDKAL--LRRLM 296
 DB 254 SNGIEGPQDRH-----SHLIADFPAYADVPVIEVEAKGKEEAIALRLM 299

D	b		125	- - - - -	TLLKQTVLST	- - - - -	KGLTERQAkyRPL	- - - - -	IMDRNACI	155													
				:	:	:	:	:	:	:													
Q	y		152	IHIGGAYG-	- - - - -	NKTATAQHONIKQLPQBIK	- - - - -	ERMtLENDdkTY	- - - - -	TTEE	196												
				:	:	:	:	:	:	:													
D	b		156	INLIvYGLSLQELVSLNMNSHIQFARNILIVPGENKLT	TRSVLSLTMETDQOLKYKYTTIPE	215																	
Q	y		197	TLQVCEQEEDVP-	-VFDFHH-	- - - - -	EYANPDHDADLNVALPRMIKTWERIGLPKPVHLSS	249															
				:	:	:	:	:	:	:													
D	b		216	PVRPOOHTNNPLUFVAFDNRGTFRWVTENDAPKALSEVATOKMIR	- - - - -	LEV	263																
Q	y		250	PKXEQAIRSHADYVDANFLERFRQWGNTIDFMIEAKOKDKALLR	294																		
				:	:	:	:	:	:	:													
D	b		264	KRAELDRRTISAQQMRMTFILRLIQGVTEKDVLVSMGMFKTKISLK	308																		
R	SULT	10																					
Q	9P977																						
I	C	Q9P977	PRELIMINARY;		PRT;		1690 AA.																
D	T	01-OCT-2000	(TReMBLrel. 15, Created)																				
D	T	01-OCT-2000	(TReMBLrel. 15, Last sequence update)																				
D	E	01-JUN-2001	(TReMBLrel. 17, Last annotation update)																				
D	E	CYRI.																					
D	E	CYRI.																					
O	S	Candida albicans (Yeast).																					
O	C	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;																					
O	C	Saccharomycetales; mitosporic Saccharomycetales; Candida.																					
X	N	NCBI_TaxID=5476;																					
O	R	[1]																					
R	A	SEQUENCE FROM N.A.																					
R	A	Ono N., Sudoh M.;																					
R	T	"Candida albicans CYRI gene."																					
R	L	Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.																					
R	L	EMBL; AB034965; BAA93553.1; -.																					
D	R	InterPro; IPR001054; Guanylit_cyclase.																					
D	R	InterPro; IPR001611; LRR.																					
D	R	InterPro; IPR003592; LRR typ.																					
D	R	InterPro; IPR003591; LRR typ.																					
D	R	InterPro; IPR003589; PP2C_catalytic.																					
D	R	InterPro; IPR001932; PP2C_domain.																					
D	R	InterPro; IPR003588; PP2C_sig.																					
D	R	InterPro; IPR000159; RA.																					
D	R	Pfam; PF00211; guanylate_cyc: 1.																					
D	R	Pfam; PF00560; LRR: 13.																					
D	R	Pfam; PF00481; PP2C: 2.																					
D	R	PRINTS; PR00019; LEURICHRPT.																					
D	R	SMART; SM00044; Cycc; 1.</																					

Query Match 6.1%; Score 101.5; DB 3; Length 1690;
Best Local Similarity 20.4%; Pred. No. 17;
Matches 66; Conservative 36; Mismatches 108; Indels 113; Gaps 15;

QY 3 FRGFSVNSLWDASPAKTLFARYSKLSYKTERKEALLTVTKANLRN-----TWRTLHY 57
DB 157 FRSSFDNSLSSSPPKD-----KVSVDVTSVSSSTVTASMSNPTISIDLNDXHD 212
QY 58 IIGH-----GIPLRFSSSIVPLATH-----PDVMDVTPFOKE-----FREIGE- 98
DB 213 IIKSPETPTAGLPTQKAERKASPTAINKWAPE-SWDVKAPIKKEPHAPKIEVAEN 271
QY 99 -----LVKTHQ-----LRTSFHPNQFTLTSPKESVTKNA 128
DB 272 DVAIDNVLEKRLPVLYGTHQVPHVNTSKDIKSSHIIRVFKEDNTFTTLCPLETTTS-- 329
QY 129 VTDMAHYRMLEAMGLADR-----SVINIHIGGAYGNKDTATQAQFHNKQLPOEIKERM 183
DB 330 -----ELLAIVOKKFFLESTTNQLSVCIGNCVKVLDEDFE-----PLKIOMGL 373
QY 184 TLENDDKTYTTEETLOVCEODVFPVDFHHFYANPDHDLNVALPRMIKTWERIGLOP 243
DB 374 LLLSG---YTEEDKRLMGLREDLSFVCKF-----VVENIFLRS 408
QY 244 KVLSSPKSEQAIRSHADYVDAN 266
DB 409 LTH-----DEEVLLSR-NYVDVN 425

RESULT 14
P95930 PRELIMINARY; PRT; 283 AA.
AC P95930;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ORF C01032.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=DSM 1617 / P2;
RX MEDLINE=97055432; PubMed=8899719;
RA Sensen C.W., Klenk H.-P., Singh R.K., Allard G., Chan C.C.-Y.,
Liu Q.Y., Penny S.L., Young F., Schenk M.E., Gaasterland T.,
Doolittle W.F., Ragan M.A., Charlebois R.L.;
RT "Organizational characteristics and information content of an archaeal
genome: 156 kb of sequence from Sulfolobus solfataricus P2.";
RL Mol. Microbiol. 22:175-191(1996).
DR EMBL; Y08256; CAA69434.1;
DR InterPro; IPR001719; AP_endonuclease_2.
DR Pfam; PF01261; AP_endonuclease2; 1.
DR SMART; SM00518; AP2EC; 1.
DR PROSITE; PS00729; AP_NUCLEASE_F2.1; UNKNOWN_1.
SQ SEQUENCE 283 AA; 32015 MW; DBFB9C13F0B5975 CRC64;

Query Match 6.1%; Score 101; DB 1; Length 283;
Best Local Similarity 21.6%; Pred. No. 1.6;
Matches 52; Conservative 48; Mismatches 87; Indels 54; Gaps 13;

QY 95 EIGELVKTQLRTSFH-PNQFTLTSPKESV-TKNVAVTDMAYHYRMLEAMGLADRSVIN 151
DB 59 ETGVARELGVRLSVHAPYFYNLCSEKEKEIEASKQRILOTDRAELMGADATIA----- 112
QY 152 IHIGGAYGNKDTATQAQFHNKQLPOEIKERM-----TLENDDKTYTTEETL 198
DB 113 IHI--AFYKWT-PECCQNVKGELEVIDKAREMGIRNVKFGVETMAKETAFGTLDVI 169
QY 199 QVCEQED--VPFVDFHHFYANPDHDLNVALPRMIKTWERIGLQPKVHLLSSPKSEQAI 256
DB 170 SISKEIKGVIPYI-DWAHTFARQGGEDYDGKIIDRLIK---ELGL---THINS-HFESLV 221

QY 257 RSHADYVD-----ANFLLERFQWGTNIDFIEAKOKDKKALLRLMDELSS 301
DB 222 YRRGKYVDEHIPIDANAPPPPEPLAKELKR-----DISTLICESPELDRALKMKEVLER 277
QY 302 I 302
DB 278 L 278

RESULT 15
Q9KB81 PRELIMINARY; PRT; 823 AA.
ID Q9KB81
AC Q9KB81;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE BH2047 PROTEIN.
GN BH2047.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001514; BAB05766.1;
KW Complete proteome.
SQ SEQUENCE 823 AA; 97055 MW; A661C8566FE8EDC7 CRC64;

Query Match 6.0%; Score 100; DB 2; Length 823;
Best Local Similarity 23.1%; Pred. No. 8.5;
Matches 75; Conservative 44; Mismatches 101; Indels 104; Gaps 18;

QY 30 KLSKTERK-----EALLTVTKANLRNMTLHYIIGHGIPHYRFSSSIVPLATHPDV-MW 83
DB 332 KLRLLKXKLLILIEITPLSITPDLSNELRS--WILK-----IKDVEHPDLSIW 377
QY 84 DFTVPQKFEFRELGEVLVKTQLRTSFH--PNQFTLTSPKESVTKNAVTDMAHYRMLEA 141
DB 378 QERLEWLWKVQAYPYMTLGQKQELFKLESTFTETWGTGQP-----RRLGG 423
QY 142 MGIADRSVI-----NIHIGGAYGN-----KDTATAQFH--QNIQK--LPQETK 180
DB 424 EIVADRNLLYECHGPLNNIKIGSIKHKVLKQDVPKWLISICAKHGQRKQEQALAEIF 483
QY 181 ERTMLENDKTYTTEETLOVCEODVFPVDFHHFYANPDHDLNVALPRMIKTWERIG 240
DB 484 KMN-----YPNEDS-----VPFLKFVHDLSNHDPVH-----TWKRW 515
QY 241 LQPKVHL-----SSPKSEQAIRSHADYVDANFLLERFQWGTNIDFIEAKOKDKA 291
DB 516 QSIKTEIEAITHAVASNPKS--VVHLSLDYQD----FDRDMAILTSPDLMI-AKKDDDS 568
QY 292 LLRLMDELSSIRGVKRRIGGCALOW 315
DB 569 YQVILGE--IHDITNMGWALQF 589

Search completed: January 15, 2002, 14:06:28
Job time: 967 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:51:16 ; Search time 57.72 Seconds
(without alignments)
226.125 Million cell updates/sec

Title: US-09-724-296-38

Perfect score: 3055

Sequence: 1 MCTTGLSLSLGRGAAPTV.....HKLKHYHDWLWENHEKLSLS 580

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA.*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*

5: /cgn2_6/ptodata/2/iaa/PCRTUS_COMB.pep.*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	146.5	4.8	1579	3	US-08-755-587-184
2	145	4.7	3418	2	US-08-639-501-2
3	145	4.7	3418	2	US-08-603-753D-4
4	145	4.7	3418	3	US-09-044-946-2
5	145	4.7	3418	3	US-08-755-587-44
6	145	4.7	3418	3	US-09-044-908-2
7	145	4.7	3418	4	US-09-099-753-4
8	145	4.7	3418	4	US-08-986-106-4
9	144	4.7	2329	3	US-08-755-587-16
10	131	4.3	754	4	US-09-214-564-2
11	129	4.2	1683	3	US-08-755-587-183
12	117	3.8	1471	3	US-08-755-587-188
13	116.5	3.8	885	1	US-08-484-105-14
14	116.5	3.8	885	1	US-08-484-106-14
15	116.5	3.8	1535	3	US-08-755-587-185
16	116.5	3.8	2343	4	US-09-324-867-2
17	115	3.8	1013	4	US-09-415-522-8
18	114.5	3.7	1589	3	US-08-755-587-189
19	113.5	3.7	1164	4	US-08-923-992A-6
20	111.5	3.6	1128	4	US-08-923-992A-6
21	111	3.6	1005	1	US-08-089-986-3
22	111	3.6	1005	1	US-08-478-585-3
23	111	3.6	1005	1	US-08-717-312-3
24	111	3.6	1005	2	US-08-266-408-3
25	111	3.6	1005	5	PCT-US94-07886-3
26	108	3.5	926	3	US-08-755-587-187
27	106	3.5	976	4	US-09-104-324B-4

28	106	3.5	1780	1	US-08-769-309A-5	Sequence 5, Appli
29	106	3.5	1780	3	US-08-994-570-5	Sequence 5, Appli
30	105.5	3.5	1098	4	US-08-923-992A-8	Sequence 8, Appli
31	105	3.4	1494	3	US-08-755-587-186	Sequence 186, App
32	105	3.4	1863	1	US-08-425-061-16	Sequence 16, Appli
33	105	3.4	1863	2	US-08-825-886-16	Sequence 16, Appli
34	104	3.4	1863	2	US-08-603-753D-2	Sequence 2, Appli
35	104	3.4	1863	4	US-09-099-753-2	Sequence 2, Appli
36	104	3.4	1863	4	US-08-986-106-2	Sequence 3, Appli
37	104	3.4	2465	2	US-08-596-291-3	Sequence 3, Appli
38	104	3.4	2465	3	US-09-100-804-3	Sequence 3, Appli
39	103.5	3.4	914	1	US-08-484-105-2	Sequence 2, Appli
40	103.5	3.4	914	1	US-08-484-106-2	Sequence 2, Appli
41	103.5	3.4	1075	2	US-08-993-228-19	Sequence 19, Appli
42	103.5	3.4	1435	2	US-08-568-459A-4	Sequence 4, Appli
43	103.5	3.4	1435	2	US-08-487-826B-4	Sequence 4, Appli
44	103	3.4	2482	1	US-08-328-254-6	Sequence 6, Appli
45	103	3.4	3248	1	US-08-353-700-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-08-755-587-184

; Sequence 184, Application US/08755587

; Patent No. 6045957

; GENERAL INFORMATION:

; APPLICANT: Futreal, Phillip A

; APPLICANT: Wooster, Richard F

; APPLICANT: Ashworth, Alan

; APPLICANT: Stratton, Michael R

; TITLE OF INVENTION: Materials and methods relating to the

; TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer

; NUMBER OF SEQUENCES: 222

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Bell Seitzer Park & Gibson

; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107

; CITY: Raleigh

; STATE: NC

; COUNTRY: USA

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/755,587

; FILING DATE: 25-NOV-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 9523959.6

; FILING DATE: 23-NOV-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 9525555.0

; FILING DATE: 14-DEC-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 9617961.9

; FILING DATE: 28-AUG-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Kenneth D Sibley

; REGISTRATION NUMBER: 31,665

; REFERENCE/DOCKET NUMBER: 5405-135

; INFORMATION FOR SEQ ID NO: 184:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1579 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; US-08-755-587-184

Query Match

Best Local Similarity 4.8%; Score 146.5; DB 3; Length 1579;

21.1%; Pred. No. 5.8e-05;

us-09-724-296-38.ra1

Tue Jan 15 14:12:16 2002

OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/755,587
 FILING DATE: 25-NOV-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB 9523959.6
 FILING DATE: 23-NOV-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB 9525555.0
 FILING DATE: 14-DEC-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB 9617961.9
 FILING DATE: 28-AUG-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Kenneth D Sibley
 REGISTRATION NUMBER: 31,665
 REFERENCE/DOCKET NUMBER: 5405-135
 INFORMATION FOR SEQ ID NO: 44:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3418 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-755-587-44

Query Match 4.7%; Score 145; DB 3; Length 3418;
 Best Local Similarity 20.9%; Pred. No. 0.0003;
 Matches 96; Conservative 67; Mismatches 158; Indels 138; Gaps 23;

QY 36 KEDVAMELERVGEDEEOMIKRSECNPLLOEPIASAOFGATAGTECRKSVPCGWERVVK 95
 DB 972 KSDISLNIDKIPEKNDYMNKAG-----LLGPISNHSFGGSFRTASNKEIKLSEHNKK 1026
 QY 96 QRLFGKTAGRFVDYFISPGILKFRSKSLANLYLHKNGETSL-KPEDDFDTVLKRGIGKSR 154
 DB 1027 SKMFFK-----DIEEQYPTSL---ACVEIVNTLALDNOKLKSQPSIN-TVSAHLQSSVV 1077
 QY 155 YKDCSMAALTSHL--ONQSNNSNWLRTSRCKKDVFPMPSSSELOESRGLSNFTSTHL 212
 DB 1078 VSDCKNSHITPQMLFSKODFNSHNHL-----TPSOKAEITE-----LSTIL 1118
 QY 213 LLKEDGVDVDFNFRKVRK-----PKGKVTILK----- 239
 DB 1119 -----ESGSOFEFTQPKRPSYILQKSTFVPEPNQMTILKTTSECRDADLHVIMNAPSIG 1174
 QY 240 -----GPIPK-----TKGCRKSCGFGVSDSKRESVCNKADAEPVAKQSOL 284
 DB 1175 QVDSKQFEGTVEIKRKFAGLKNDCNKSASGYLTDE-----NEVGRFGYSAHGTL 1227
 QY 285 D-RVTCISDAGACGETLSVTSENSLVKKKRSLSGSGSNFCSEQKTSIINKFCSAKDSE 343
 DB 1228 NVSTEALQKAVKLFSDIENISEETS---AEVHPISLSSSKCHD---SVVSMF---KLEN 1277
 QY 344 HNEKYEDTFLESEE-----IGTKVEVVERKEHLHTDILKRGSEMDNN--CSP 388
 DB 1278 HNDK---TVSEKNNKCOLIQNNNIEMTGTFFVEI-----TENYKRNTEDEKNKYTAA 1327
 QY 389 TRK---DFTGEKIFQEDTIPRTQIERRTKTSLYFSSKYN 423
 DB 1328 SRNSHNLEFDGSDSSKNDIV---CIHKDETDLFLTDOHN 1363

RESULT 6

US-09-044-908-2
 ; Sequence 2, Application US/09044908
 ; Patent No. 6124104
 ; GENERAL INFORMATION:
 ; APPLICANT: Tavtligian, Sean V.
 ; APPLICANT: Kamb, Alexander
 ; APPLICANT: Simard, Jacques
 ; APPLICANT: Couch, Fergus

REGISTRATION NUMBER: 28,957
 REFERENCE/DOCKET NUMBER: 24884-116802-04
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-962-4810
 TELEFAX: 202-962-8300
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3418 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-044-946-2

Query Match 4.7%; Score 145; DB 3; Length 3418;
 Best Local Similarity 20.9%; Pred. No. 0.0003;
 Matches 96; Conservative 67; Mismatches 158; Indels 138; Gaps 23;

QY 36 KEDVAMELERVGEDEEOMIKRSECNPLLOEPIASAOFGATAGTECRKSVPCGWERVVK 95
 DB 972 KSDISLNIDKIPEKNDYMNKAG-----LLGPISNHSFGGSFRTASNKEIKLSEHNKK 1026
 QY 96 QRLFGKTAGRFVDYFISPGILKFRSKSLANLYLHKNGETSL-KPEDDFDTVLKRGIGKSR 154
 DB 1027 SKMFFK-----DIEEQYPTSL---ACVEIVNTLALDNOKLKSQPSIN-TVSAHLQSSVV 1077
 QY 155 YKDCSMAALTSHL--ONQSNNSNWLRTSRCKKDVFPMPSSSELOESRGLSNFTSTHL 212
 DB 1078 VSDCKNSHITPQMLFSKODFNSHNHL-----TPSOKAEITE-----LSTIL 1118
 QY 213 LLKEDGVDVDFNFRKVRK-----PKGKVTILK----- 239
 DB 1119 -----ESGSOFEFTQPKRPSYILQKSTFVPEPNQMTILKTTSECRDADLHVIMNAPSIG 1174
 QY 240 -----GPIPK-----TKGCRKSCGFGVSDSKRESVCNKADAEPVAKQSOL 284
 DB 1175 QVDSKQFEGTVEIKRKFAGLKNDCNKSASGYLTDE-----NEVGRFGYSAHGTL 1227
 QY 285 D-RVTCISDAGACGETLSVTSENSLVKKKRSLSGSGSNFCSEQKTSIINKFCSAKDSE 343
 DB 1228 NVSTEALQKAVKLFSDIENISEETS---AEVHPISLSSSKCHD---SVVSMF---KLEN 1277
 QY 344 HNEKYEDTFLESEE-----IGTKVEVVERKEHLHTDILKRGSEMDNN--CSP 388
 DB 1278 HNDK---TVSEKNNKCOLIQNNNIEMTGTFFVEI-----TENYKRNTEDEKNKYTAA 1327
 QY 389 TRK---DFTGEKIFQEDTIPRTQIERRTKTSLYFSSKYN 423
 DB 1328 SRNSHNLEFDGSDSSKNDIV---CIHKDETDLFLTDOHN 1363

RESULT 5

US-08-755-587-44
 ; Sequence 44, Application US/08755587
 ; Patent No. 6045997
 ; GENERAL INFORMATION:
 ; APPLICANT: Futreal, Phillip A
 ; APPLICANT: Wooster, Richard F
 ; APPLICANT: Ashworth, Alan
 ; APPLICANT: Stratton, Michael R
 ; TITLE OF INVENTION: Materials and methods relating to the
 ; TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer
 ; TITLE OF INVENTION: susceptibility gene and uses thereof.
 ; NUMBER OF SEQUENCES: 222
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Bell Seltzer Park & Gibson
 ; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
 ; CITY: Raleigh
 ; STATE: NC
 ; COUNTRY: USA
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible

APPLICANT: Rommens, Johanna
 APPLICANT: Weber, Barbara
 TITLE OF INVENTION: Chromosome 13-Linked Breast Cancer
 NUMBER OF SEQUENCES: 124
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Venable, Baetjer, Howard & Civiletti
 STREET: 1201 New York Avenue N.W., Suite 1001
 CITY: Washington
 STATE: DC
 COUNTRY: USA
 ZIP: 22204
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/044,908
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/639,501
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/576,559
 FILING DATE: 21-DEC-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/575,359
 FILING DATE: 20-DEC-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/573,779
 FILING DATE: 18-DEC-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Ihnen, Jeffrey L.
 REGISTRATION NUMBER: 28,957
 REFERENCE/DOCKET NUMBER: 24884-116802-04
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-962-4810
 TELEFAX: 202-962-8300
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3418 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-044-908-2

Query Match 4.7%; Score 145; DB 3; Length 3418;
 Best Local Similarity 20.9%; Pred No. 0.0003;
 Matches 96; Conservative 67; Mismatches 158; Indels 138; Gaps 23:
 QY 36 KEDVAMELERVGEDEOMIKRSECNPLLOEPIASQAGTAGTECRKSVPCGWERVVK 95
 Db 972 KSDLSLNDIKPEKNDYNNKAG-----LLGPISNHSFGGSRFTASNKEIKLSHNKK 1026
 QY 96 QRLFGKTAGRFQVYFTSPQGLFKRSKSLANYLHNGETSL-KPEDFDFTVLSKRGIKSR 154
 Db 1027 SKWFFK-----DLEEQYPTSL---ACVEIVNTALDNDOKLKSPOSIN-TVSAHLQSSV 1077
 QY 155 YKDCSMAALTSL--ONQSNNSNWLRTSRCKCKDVFMPPSSSELSQESRGLSNFTSTHL 212
 Db 1078 VSDCKNSHITPQMLFSKQDFNSNHL-----TPSQAEITE-----LSTIL 1118
 QY 213 LLKEDGVDVNVFRVKR-----PKGKVITLK----- 239
 Db 1119 -----EESGQFETFRKPSYILQKSTFEVPEQNQMTILKTTSEECRADLHVINNAPSIG 1174
 QY 240 -----GPIPKK-----TKGCRKSCGFGVQSDSKRESVCKADASEPVAQKSQ 284
 Db 1175 QVDSSKQFEGTVEIKRKAGLLKNDCKNSASGYLTDE-----NEVGFRGFYSAHGTLK 1227

QY 285 D-RTVCISDAGAGETLSVTSEENSLVKKKRSLSGSGNFCSEQKTSGLINKFCSAKDSE 343
 Db 1228 NVSTEALQKAVKLFSDIENISEETS---AEVHPISLSKSKCHD---SVVSMF---KLEN 1277
 QY 344 HNEKYEDTFLSEEE-----IGTKVEVVVERKEHLHTDILKRGSEMDNN--CSP 388
 Db 1278 HNDK---TVSEKNNKCOLLQNNIEMTTGTFFVEE-----TENYKRNTEDEKNTAA 1327
 QY 389 TRK----DETGEKIFQEDTIPRTOHERRKTSLYFSSKYN 423
 Db 1328 SRSHNLEFDGSDSSKNQTV---CIHKDETLLFTDQHN 1363
 RESULT 7
 US-09-099-753-4
 : Sequence 4, Application US/09099753
 : Patent No. 6149903
 : GENERAL INFORMATION:
 : APPLICANT: HOLT, JEFFREY T.
 : APPLICANT: JENSEN, ROY A.
 : APPLICANT: PAGE, DAVID L.
 : APPLICANT: KING, MARY-CLAIRE
 : APPLICANT: SZABO, CSILLA I.
 : APPLICANT: JETTON, THOMAS L.
 : APPLICANT: ROBINSON-BENION, CHERYL L.
 : APPLICANT: THOMPSON, MARILYN E.
 : TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2
 : TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON
 : TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.
 : NUMBER OF SEQUENCES: 29
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: ARLES A. TAYLOR, JR.
 : STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
 : STREET: BOULEVARD
 : CITY: DURHAM
 : STATE: NORTH CAROLINA
 : COUNTRY: USA
 : ZIP: 27707
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage
 : COMPUTER: IBM PC/XT/AT compatible
 : OPERATING SYSTEM: Windows 3.1
 : SOFTWARE: WORD PERFECT 6.1 and ASCII
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/099,753
 : FILING DATE:
 : CLASSIFICATION:
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 08/603,753
 : FILING DATE: 20 FEB 1996
 : APPLICATION NUMBER: U.S. 08/373,799
 : FILING DATE: 17 JAN 1995
 : ATTORNEY/AGENT INFORMATION:
 : NAME: ARLES A. TAYLOR, JR.
 : REGISTRATION NUMBER: 39,395
 : REFERENCE/DOCKET NUMBER: 1242/2
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (919) 493-8000
 : TELEFAX: (919) 419-0383
 : TELEX:
 : INFORMATION FOR SEQ ID NO: 4:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 3418
 : TYPE: amino acid
 : STRANDEDNESS: single
 : TOPOLOGY: unknown
 : MOLECULE TYPE: protein
 : HYPOTHETICAL: no
 : ANTI-SENSE: no
 : ORIGINAL SOURCE:
 : ORGANISM: Homo sapiens sapiens
 : INDIVIDUAL ISOLATE:
 : DEVELOPMENTAL STAGE: adult

TISSUE TYPE: female breast
CELL TYPE: normal breast tissue
CELL LINE: HMEC
ORGANELLE: no

FEATURE:
NAME/KEY: BRCA2 protein
LOCATION: 1 to 3418; Genbank locus HSU43746
IDENTIFICATION METHOD:
OTHER INFORMATION: BRCA2 protein has a negative
OTHER INFORMATION: regulatory effect on growth of human mammary cells.
PUBLICATION INFORMATION:
AUTHORS: Wooster, R. et al.
TITLE: Identification of the breast cancer
TITLE: susceptibility gene BRCA2
JOURNAL: Nature
VOLUME: 379
PAGE: 789-792
DATE: 1995
RELEVANT RESIDUES IN SEQ ID NO: 4: granin box
RELEVANT RESIDUES IN SEQ ID NO: domain at amino acids 3334-3344

US-09-099-753-4

Query Match 4.7%; Score 145; DB 4; Length 3418;
Best Local Similarity 20.9%; Pred. No. 0.0003;
Matches 96; Conservative 67; Mismatches 158; Indels 138; Gaps 23;

QY 36 KDVAMELERYGEDEQWMTKRSECNPLQLQEPISAAQFCATAGTECRKSVPCGWVVK 95
DB 972 KDSIDSLNDIKIPKNNNDYMKWAG-----LLGPISNHSFGGSPRTASKEIKLSEHNKK 1026
QY 96 QRLFGKTAGRDVDFVPSIQGLKFRKSSLANYLHKNGETSL-KPEDFDFTVLSKRGIKSR 154
DB 1027 SKMFKE-----DIERQYPTSL--ACVEIVNTALDNQKKLSQPQIN-TVSAHLQSSVV 1077
QY 155 YKDCSMAALTSHL--QNGSNNSNNLRTSRCKKDKVMPSPSSSELOESRGLSNFTSTHL 212
DB 1078 VSDCKNSHITPWLFSKODFNSNHL-----TPSQAEITE-----LSTIL 1118
QY 213 LKDEGDDVDVNFVRVK-----PKGKVITLK----- 239
DB 1119 ----EESQGFETQFRKPSYILOKSTFEVPEQNMILKTSTSECRDADLHVTMNPSTG 1174
QY 240 -----GPIPK-----TKGCRKSCSGFVQSDSKRSCVKNKADAESFVAQKSQL 284
DB 1175 QVDSQKQFGTVEIKRFAKGLKNDCKNSAGYLTD-----NEVGRFGFYSAHGTKL 1227
QY 285 D-RTVCISDAGAGETVLSYTSSENSLVKKKRSLSGSSNFCSCSEKTKSGIINKFCSAKDSE 343
DB 1228 NVSTEALQKAVKLFSIDIENISETS---AEVHPISLSKKCHD---SVYSMF---KIEN 1277
QY 344 HNKEDYDTFLEEE-----IGTKVEVVERKEHLHTDILKRGSEDMNN--CSP 388
DB 1278 HNDK--TVSEKNKKCOLIQNNIEMTGTFFVEI-----TENYKRETNEDNKYTAA 1327
QY 389 TRK----DFTGKIFEDIPRTQIERRKTSLYFSKYN 423
DB 1328 SRNSHNLEFDGSDSKNDIV--CIHKDETLLFTDOHN 1363

RESULT 8
US-08-986-106-4
; Sequence 4, Application US/08986106
; Patent No. 6177410
; GENERAL INFORMATION:
; APPLICANT: HOLT, JEFFREY T.
; APPLICANT: JENSEN, ROY A.
; APPLICANT: KING, MARY-CLAIRE S.
; APPLICANT: STEINER, MITCHELL S.
; APPLICANT: ROBINSON-BENION, CHERYL L.
; APPLICANT: THOMPSON, MARILYN E.
; TITLE OF INVENTION: THERAPEUTIC METHODS FOR
; TITLE OF INVENTION: PROSTATE CANCER

QY 285 D-RTVCISDAGACGETLSVTSEENSLVKKKERSLSSGNSFCSEBQKTSGLINKFCSAKDSE 343
 Db 1228 NVSTEALQKAVKLFSDIENISEETS---AEVHPISLSKCHD---SVVSMF---KIEN 1277
 QY 344 HNEKYEDTFLESEE-----IGTKVEVVERKEHLHTDILKRGSEMDNN--CSP 388
 Db 1278 HNDK---TVSEKNKCOLILONNIEMTTCTFVEEI-----TENYKNTENEDNKYTAA 1327
 QY 389 TRK---DFTGEKIFQEDTIPRTQIERKTSLYFSSKYN 423
 Db 1328 SRNSHNLFFDGSDDSKNDTV---CIHKDETLLFTDOHN 1363

RESULT 9

US-08-755-587-16
 ; Sequence 16, Application US/08755587
 ; Patent No. 6045997

GENERAL INFORMATION:

; APPLICANT: Futreal, Phillip A
 ; APPLICANT: Wooster, Richard F
 ; APPLICANT: Ashworth, Alan
 ; APPLICANT: Stratton, Michael R
 ; TITLE OF INVENTION: Materials and methods relating to the
 ; TITLE OF INVENTION: Identification and sequencing of the BRCA2 cancer
 ; TITLE OF INVENTION: susceptibility gene and uses thereof.
 ; NUMBER OF SEQUENCES: 222
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Bell Seltzer Park & Gibson
 ; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
 ; CITY: Raleigh
 ; STATE: NC
 ; COUNTRY: USA

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/755,587
 ; FILING DATE: 25-NOV-1996

PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 9523959.6

FILING DATE: 23-NOV-1995

; APPLICATION NUMBER: GB 9525555.0

FILING DATE: 14-DEC-1995

; APPLICATION NUMBER: GB 9617961.9

FILING DATE: 28-AUG-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Kenneth D Sibley

; REGISTRATION NUMBER: 31,665

; REFERENCE/DOCKET NUMBER: 5405-135

; INFORMATION FOR SEQ ID NO: 16:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2329 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-755-587-16

Query Match 4.7%; Score 144; DB 3; Length 2329;
 Best Local Similarity 21.1%; Pred. No. 0.0002;
 Matches 97; Conservative 66; Mismatches 150; Indels 138; Gaps 24;

QY 36 KEDVAMELVRGDEEOMITKRSECNPLQBPPIASAOFCATAGTECRKSVPCGWERVVK 95

Db 691 KSDISLNDIKPEKNDYMKWAG-----LLGPISNHSEGGSFRTASKEIKLSEHNKK 745

QY 96 QRLFGKTAGFDVYFYSPOGLFKRSKSIANYLHKNGETSL-KPEDFDFTVLSKRGIKSR 154

Db 746 SKMFFK-----DIEEQYPTSL---ACVEIVNTLALDNOKLKSQPSIN-TYSAHLQSSV 796

QY 155 YKDCSMAALTSHL--QNQSNNSNNLRTSKCKKQVMPSPSSSELOESRGLSNFTSTHL 212
 Db 797 VSCKKNSHIITPOMLFKQDFNSHNHL-----TPSQEQITE-----LSTIL 837
 QY 213 LLKEDGCVDDVNRKVRK-----PKGKVTILK----- 239
 Db 838 ---EDSG-SQFEFTQFRKPSYILOKSTFEVPENQMTILKTTSSECRDADLHVIMNAPSIG 893
 QY 240 -----GPIKK-----TKGCRKSCSGFVOSDKRSVCKADAESAEPVAAKSQL 284
 Db 894 QVDSSKQFEGTVEIKRFAGALLKNCNKSASGLTDE-----NEVGRFGYSAHGTKL 946
 QY 285 D-RTVCISDAGACGETLSVTSEENSLVKKKERSLSSGNSFCSEBQKTSGLINKFCSAKDSE 343
 Db 947 NVSTEALQKAVKLFSDIENISEETS---AEVHPISLSKCHD---SVVSMF---KIEN 996
 QY 344 HNEKYEDTFLESEE-----IGTKVEVVERKEHLHTDILKRGSEMDNN--CSP 388
 Db 997 HNDK---TVSEKNKCOLILONNIEMTTCTFVEEI-----TENYKNTENEDNKYTAA 1046
 QY 389 TRK---DFTGEKIFQEDTIPRTQIERKTSLYFSSKYN 423
 Db 1047 SRNSHNLFFDGSDDSKNDTV---CIHKDETLLFTDOHN 1082

RESULT 10

US-09-214-564A-2

; Sequence 2, Application US/09214564A

; Patent No. 6150515

; GENERAL INFORMATION:

; APPLICANT: Sharp, Phillip A.

; APPLICANT: Zhou, Qiang

; TITLE OF INVENTION: TAT-SF: Cofactor For Stimulation Of Transcriptional

; FILE REFERENCE: M0656/7042

; CURRENT APPLICATION NUMBER: US/09/214,564A

; PRIOR FILING DATE: 1999-08-18

; PRIOR FILING DATE: 1996-07-03

; PRIOR FILING DATE: 1996-12-13

; PRIOR FILING DATE: 1997-07-03

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 754

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-214-564A-2

Query Match 4.3%; Score 131; DB 4; Length 754;
 Best Local Similarity 20.3%; Pred. No. 0.00057;
 Matches 88; Conservative 69; Mismatches 161; Indels 116; Gaps 17;

QY 13 DRGAAPTIVTSERLV-----PDP-----NDLRKEDVAMELERVGEDEEQMMIKRS 58

Db 252 ERRAGPSRMHRHVITKNMFHMDPDDPLVLINEIR-EDLRVCSKFCQIRKLLLFDRH 310

QY 59 SECNPLQBPPIASAOFCATAGTECRKSVPCGWERVVKORLFCK---TAGRFDVYFISQ 114

Db 311 -----PDGVASVSRDDEEADYCIOTLDGRW-----EGRQITAAQWD----- 348

QY 115 GLKFRKSSIANYLHKNGETSLKPEDFDFTVLSKRGIKRSYKDCSMAAL-TSHLQNSNS 174

Db 349 -----GTTDYQVEE-----TSREEREKLRGWEAFNLNAPNAR 380

QY 175 NNWLRTSRCKKQVMPSPSSSELOQB---SRGLSNSTSTHLLKEDEGVDDVNFRR----- 227

Db 381 GUSVQILSLLRK---AGPSRARRHFSEHPSTSKMNAQETATGMAFEPIDKKFKTEKDG 437

```

QY 228 -----VRKPKGVITLKGIPKTKKCKRSCSGFVQSDSKRESVCNKADAESE 277
Db 438 EFEEGASENNAKESPEKEAEGCPEKESBEGCPK--RGFEGSCSQKESBEGPNVRGSEE 495
QY 278 VAKQSOLDRTVCISDAGACGETLSVTSEENSLVKKERSLSSGS-----NFCSEQKTS 330
Db 496 DSPKRESK-----KTLKNDCEENGLAKESDDLNKSEEEVGVTKSEEDDSE 544
QY 331 GIINKFCSAKOSEHNEKYEDTFLESEETGTVKEVVER--KEHLHTDILKRGSEMDNMCSP 388
Db 545 KESDEDCSKQSEKGESE-----REFEENGLEKDLDEEGSEKELHENVLDKRELEENDSENS 599
QY 389 TRKDFTEGKIFQED 402
Db 600 EFEDDGSSEKVLDEE 613

RESULT 11
US-08-755-587-183
; Sequence 183, Application US/08755587
; Patent No. 6045997
; GENERAL INFORMATION:
; APPLICANT: Futreal, Phillip A
; APPLICANT: Wooster, Richard F
; APPLICANT: Ashworth, Alan
; APPLICANT: Stratton, Michael R
; TITLE OF INVENTION: Materials and methods relating to the
; TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer
; TITLE OF INVENTION: susceptibility gene and uses thereof.
; NUMBER OF SEQUENCES: 222
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer Park & Gibson
; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
; CITY: Raleigh
; STATE: NC
; COUNTRY: USA
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/755,587
; FILING DATE: 25-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9523959.6
; FILING DATE: 23-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9525555.0
; FILING DATE: 14-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9617961.9
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenneth D Sibley
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5405-135
; INFORMATION FOR SEQ ID NO: 183:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1683 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-755-587-183

Query Match 4.28; Score 129; DB 3; Length 1683;
Best Local Similarity 20.48; Pred No. 0.0035;
Matches 103; Conservative 67; Mismatches 175; Indels 160; Gaps 19;

QY 11 LGDRGAAPTV-----TSSRLVDPDPNDLR--KEDVAMELEKVEDEQMMIKRSSECN-- 62
Db 297 IGDKQAQVSIKKDLDSIAVHDLAEENRNSKQHLKMTLQDLDKLSIDLSLNXKSGDND 356

```

```

QY 63 -----PLLQEPFLASQFAGTAGTECRKSVPCGWVERVVQKRLFCKTAGREDVYFISPOGL 117
Db 357 YMDKAGLLDPLSNHISFGSGFRASNKEIKLSEHNKSKMFKDIE--EQYPTSLACVE 414
QY 118 FRSKSLANYLHKNGETSLKPDFTVLKSRGKIKRYKDCSMAALTSHLN----- 169
Db 415 JVNLTALAN-----OKLSKXXXXDPOINTVSAHLQSSVVVSDCE 454
QY 170 -----QSNNSNNLRTSRCKKDVFMPPSPSSSELOESRGLSNFTSTHLLKED 217
Db 455 DSHTAPQMLSKQDFNSNHL-----TPSOKAEITE-----LSTIL---E 491
QY 218 EGVDVNFVRK-----PKGKVTILK----- 239
Db 492 ESGSQFETQFKPKSHIAQKNTFEVPEPNQMTILNTSEWKDADLHLIVNAPSIOVDSSK 551
QY 240 -----GPIPK---KTKGCRKSCSGFVQSDSKRE-----SYCNKADAESEPVAKOSOL 284
Db 552 QFEGSAGIKQKFAKLLKSSCNKASGYLTDENEVEFRGFGYSALGTCLNVSSSEALOKAVKL 611
QY 285 DRTVCISDAGACGETLSVTSEENSLVKKERSLSSGSSENFCEQKTSGLINKFCSEKADSEH 344
Db 612 -----FSDIENLSEETSAEVDPISSLSSSKYHD-SVASMFKLENQXD-----KSSEK 657
QY 345 NEKYEDTFLESEETGTVKEVVERKEHLHTDILKRGSEMDN-----CSPTRKDFTEGK 397
Db 658 NNKQQLILQNNLEMTGIFVEE-----NTENYKRNTEDEKNTGASRNSXANLEXSDGSD 712
QY 398 IFQEDTIPRTQIERBKTSLYTSSKY 422
Db 713 SSKNDTV---YIHKDETDLPFIDQH 734

RESULT 12
US-08-755-587-188
; Sequence 188, Application US/08755587
; Patent No. 6045997
; GENERAL INFORMATION:
; APPLICANT: Futreal, Phillip A
; APPLICANT: Wooster, Richard F
; APPLICANT: Ashworth, Alan
; APPLICANT: Stratton, Michael R
; TITLE OF INVENTION: Materials and methods relating to the
; TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer
; TITLE OF INVENTION: susceptibility gene and uses thereof.
; NUMBER OF SEQUENCES: 222
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer Park & Gibson
; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
; CITY: Raleigh
; STATE: NC
; COUNTRY: USA
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/755,587
; FILING DATE: 25-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9523959.6
; FILING DATE: 23-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9525555.0
; FILING DATE: 14-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9617961.9
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenneth D Sibley
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5405-135
; INFORMATION FOR SEQ ID NO: 183:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1683 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-755-587-188

```


Search completed: January 15, 2002, 13:51:25
Job time: 354 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:52:32 ; Search time 74.94 Seconds
(without alignments)
589.555 Million cell updates/sec

Title: US-09-724-296-38
Perfect score: 3055
Sequence: 1 MGTGTGLESLSLGRGAAPTV.....HKLNKYHDLWENHEKLSLS 580

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 20000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	209	6.8	476	2 S57963	methyl CpG binding
2	201	6.6	492	2 A41907	methyl-CpG-binding
3	145	4.7	1301	2 S51323	SAC3 protein - yea
4	145	4.7	3418	1 G02334	breast cancer tumo
5	141.5	4.6	782	2 S27833	rhostry-associated
6	138.5	4.5	500	2 S55785	nucleolar protein
7	137	4.5	560	2 S53382	protein YKR028c ho
8	136.5	4.5	598	2 B40713	cylicin I - human
9	136.5	4.5	1359	2 T34036	hypothetical prote
10	135	4.4	822	2 T41622	probable ABC trans
11	132.5	4.3	286	2 A72582	probable A/G-speci
12	132	4.3	1819	2 A71928	cag island protein
13	131.5	4.3	997	2 T43523	cut17 protein - fi
14	130	4.3	1927	2 G64585	cag pathogenicity
15	128.5	4.2	946	2 A96748	hypothetical prote
16	128	4.2	853	2 T51505	hypothetical prote
17	126	4.1	561	2 H86442	unknown protein (i
18	124.5	4.1	491	2 T50346	hypothetical prote
19	124	4.1	650	2 T33350	hypothetical prote
20	124	4.1	1702	2 T14050	protein kinase (EC
21	122.5	4.0	6689	2 S55024	nebulin, skeletal
22	122	4.0	651	2 C86333	hypothetical prote
23	122	4.0	3924	2 S37431	ankyrin 2, neurona
24	120.5	3.9	911	2 S51441	hypothetical prote
25	120.5	3.9	1147	2 JN0599	DNA-binding protei
26	120.5	3.9	1233	2 S56271	hypothetical prote
27	120	3.9	1131	2 A49393	activator 1 large
28	120	3.9	1440	2 T33813	hypothetical prote
29	120	3.9	1658	2 S55101	hypothetical prote

30	120	3.9	2253	2 T30336	nuclear/mitotic ap
31	119.5	3.9	1740	2 T43773	hypothetical prote
32	119.5	3.9	3122	2 T17202	DNA-directed DNA p
33	119	3.9	646	2 F71620	hypothetical prote
34	119	3.9	797	2 H84919	hypothetical prote
35	119	3.9	991	2 H86168	hypothetical prote
36	119	3.9	1002	2 T30546	major surface glyco
37	118.5	3.9	891	2 B84614	hypothetical prote
38	118.5	3.9	1040	2 E71412	hypothetical prote
39	118	3.9	533	2 B84590	hypothetical prote
40	118	3.9	734	2 B42680	nucleolus-cytoplas
41	118	3.9	1200	2 A46194	neurofilament prot
42	118	3.9	1377	2 T51447	transcription regu
43	118	3.9	3329	2 T42205	breast cancer susc
44	118	3.9	3329	2 T30904	breast cancer tumo
45	117.5	3.8	957	2 T10633	hypothetical prote

ALIGNMENTS

RESULT 1
S57963
methyl CpG binding protein 2 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 05-Nov-1999
C:Accession: S57963
R:d'Esposito, M.; Quaderi, N.A.; Ciccodicola, A.; Bruni, P.; Esposito, T.; D'Urso, M.
Submitted to the EMBL Data Library, July 1995
A:Description: Physical mapping and expression analysis of an X-linked gene encoding
A:Reference number: S57963
A:Accession: S57963
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-476 <DES>
A:Cross-references: EMBL:X89430; NID:g899295; PIDN:CAA61599.1; PID:g899296

Query Match 6.8%; Score 209; DB 2; Length 476;
Best Local Similarity 24.0%; Pred. No. 4.1e-06;
Matches 104; Conservative 53; Mismatches 149; Indels 128; Gaps 17;

QY	37	EDVAMELERVGEDEQEMTKRSECNPLIQEPTASAQFG-----ATAGTE	81
DB	13	KDKPLFKVKVKDKKEKEGKEHPQVPSAHHSPEAEAKAETSESGSARICEASAPK	72
QY	82	CRKSV-----PCGMERVVKQRLFGKTAGRFVDYFISPOGLKFRKSSILANYL	128
DB	73	QRRSILRDGRPMYDDPTLPFGWTRKLKQKRSAGKYDVLINPQKAFRSKVELIAVF	132
QY	129	HKNGETSLKPEDFDFTVLSKRGTIKSRYSKDCSMAALTSHLQNSNNSNWLRTSRCKKDV	188
DB	133	ERVGDTSLDPNDFDFTV--TGRGSPSR-----REQKPKK--	165
QY	189	FMPGSSSELQESRGL---SNFTSTHLLKDEGVDVDFNFRKVRKPKGVITLKGIPKK	245
DB	166	---PKSPKAPGTGRGRGKSGTTPKKAATSEGVQVK--RVLEKSPGK--LLVMPPF-Q	217
QY	246	TKKGCRKSCSGFVQSDS----KREYCNKADAESEPAQKSQLDRTVCTSDAGACGETUS	301
DB	218	TSPGGKAEGGGATTSTQVMVVKRPGGRKKAADPOAIPKKR-----GRKPGSVVA	267
QY	302	VTFSEENSLVKKRSLSUSSNFCSEQKTSGITNKKFCSAKDSEHNKEYEDTFLESEIGTK	361
DB	268	AAAAEAKKAVKRESSIR-----SVQETVLPKK-----RKTRT-----VS	303
QY	362	VEVVERKEHLHTDIL--KRGSEMDNNSCPTRDTEKIFQEDTIPRTOIERRRTSLVFS	419
DB	304	IEVKVVKVPLLVSTLCEKSGKGLKTCCKSPGRK-----SKESPGR-----	344
QY	420	SKYNKEALSPRRK 433	
DB	345	---SSSASSPPPKKE 355	

A;Description: The SAC3 gene codes for a nuclear protein required for normal mitosis.

A;Reference number: S51323

A;Accession: S51323

A:Molecule type: DNA

A:Residues: 1-1301 <BAU>

A:Cross-references: EMBL:Z47805; NID:q634085; PID:q634086

R:Murphy, L.; Richards, C.; Harris, D.

submitted to the EMBL Data Library, July 1995

A:Reference number: S57971

A: Accession: S57983
A: Reference Number: S57971

A: Molecule type: DNA
Accession: 357503

A: MOLECULE TYPE: DNA
A: Residues: 1-1301 <MDR>

A, RESIDUES: I 1301 <MOK>
A:Cross-references: EMBL:Z50046: NTD:a899393: PTD:a899406: MIPS:YDR159W

A: Cross-References: EMBL: Z30040, NID: g099999, FID: g099999, MIPS: D5K1334
A: Experimental source: strain AB972

A; EXPERIMENTAL SOURCE: STRAIN AB972

R; BAUEL, A.; KOETTLING, K.
yeast 12 965-975. 1996

A. Title: Characterization of the *SAC3* gene of *Saccharomyces cerevisiae*

A; TITLE: CHARACTERIZATION OF THE SAC3 GENE OF SACCAROMYCES CEREVISIAE.
A. Reference number: S71744; MUID: 97027306

A;Reference Number: S71744; MOLD:97027306
A:Accession: S71744

A;ACCESSION: S71744
A;MOLECULE type: DNA

A; Molecule type: DNA
A-Residues: 1-1301 <RAW>

A;RESIDUES: I-130I <BAW>
A:Cross-references: FMRI:247805. NID:0634085. PIDN:CAA87767 1: PID:0634086

A; CROSS-REFERENCES: EMBL:Z47603; NID:9034083; FIDN:CA86767.1, FID:9034080
C-Genetics:

C; GENETICS:
A-Gene: SCD-SAC3- [EP1

A/Gelle: SGD:SAC3; LEPI
A-Cross-references: SGD:S0002566; MIPS:VDR159.w

A;Cross-references: SGD:S0002566; MIPS:YDR159W
A;Map position: 4P

A;Map position: 4R
C;Emulation:

C;Function: potential regulator of lowing hormone gene expression

A;description: potential regulator of leucine permease gene expression

C; keywords: nucleus; transmembrane protein

Query Match 4.7%; Score 145; DB 2; Length 1301;
Best Local Similarity 21.9%; Pred. No. 0.22;
Matches 116; Conservative 61; Mismatches 185; Indels 168; Gaps 25;

Qy	123	SLANYLHKNCETSLKPEDDFIVLSRGRKTSRYKDC-----S\$59
Db	510	TLOHYSHKISETOPLKKTY-LTCLERRLOKTTYGLINGEDNLASSVYVKDPKKDRIPS 568
Qy	160	MAALTSHLONOSNNWNLRTSKCK-----KDFMPPSSSELQESRGLNSTSTHL 212
Db	569	IADOSFLMENFQNYNEKLNQNSSVKPOINTSPKRVATREPHFPFSQESKOLPOLISOHT 628
Qy	213	LLKEDEGVDDVNRKVRKPKGVWTLILKGIPIKTKKGRKSGGVQVOSDKRESVCNKAD 272
Db	629	L-----STNPLLTPOVHGDLSEKQOQIKTIVTDG-----GSPFVFDQAGCNSTVEASK 676
Qy	273	AESEPVAKSQLDRTWCISDAGAGCTLSLTREENSELSVKKKRSLSGSGNFCSEKTSOI 332
Db	677	A-----HMITSTNGAYDEKLS-SEQEMRKKEQORI-----EEKT--- 712
Qy	333	INKFCSAKUSEHNEKYEDTFLESEETGTV--EVVERKEHLHTDILKRGSEMDNNCSPTR 390
Db	713	-----OLKKQEN---ADKOVITEQIANDLVKEVNSSV---ISIVKREFSEAN---YR 757
Qy	391	KDTGTEKIFOEDTIPR-----TOIERKTSLSYTSCKYKNEALSPRRKAKKKW----- 438
Db	758	KDFI-----DTMTRELYDAFLHERLYLIYMDSRAEKLNSTLKKKFKKQWASYSQAK 810
Qy	439	-----TTPKSPNLMVOE-----TLPHDPKWLIIATIFLNRTSGKKMAI--PV-- 477
Db	811	KNRILBEKKKEEIKLWSHQLVGPFRKKTCLFTPTPKYKGVNWSFMLSDDKNLIFSPVND 870
Qy	478	-----LWKFLB-----KYPSSAEVARTADWRDVSSELLKPLGLYDLUR 512
Db	871	EFNKPATHLTKISKLWRPEMQSIYDNLTKKEP-----SNSLTPANLFI-Y 916
Qy	513	AKTIVRFSDEYLTQKQKYPHELHGICQYKND--SYRIFCVNEKWQVHPED 560
Db	917	AKOWTSLNWNWILSKFNLO-TAODSKKFSNNIISSRIICIDD--EYEPSD 963

RESULT 2

AA1907 methyl-CpG-binding protein 2 - rat
N:Alternate names: chromosomal protein MeCP2
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 08-Oct-1999
C:Accession: AA1907; S41461
R:Lewis, J.D.; Meehan, R.R.; Henzel, W.J.; Maurer-Fogy, I.; Jeppesen, P.; Klein, F.; Bird
Cell 69, 905-914, 1992
A:Title: Purification, sequence, and cellular localization of a novel chromosomal protein
A:Reference number: AA1907; MUID:92298389
A:Accession: AA1907
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-492 <LEW>
A:Cross-references: GB:M94064; NID:g205361; PIDN:AAA41584.1; PID:g205362
R:Nan, X.; Meehan, R.R.; Bird, A.
Nucleic Acids Res. 21, 4886-4892, 1993
A:Title: Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2.
A:Reference number: S41461; MUID:94232813
A:Contents: annotation; methyl CpG-binding domain
C:Keywords: chromosomal protein; DNA binding
F:78-162/Domain: methyl-CpG-binding #status experimental <MCG>

Query Match 6.6%; Score 201; DB 2; Length 492;
Best Local Similarity 22.9%; Pred. No. 1.4e-05;
Matches 108: Conservative 57; Mismatches 156; Indels 150; Gaps 18;

[illegible]

RESULT 3
SAC3 protein - yeast (*Saccharomyces cerevisiae*)
N3: Alternate names: protein YD8358.13; protein YDR159W
C: Species: *Saccharomyces cerevisiae*
C: Date: 23-Feb-1995 #sequence_revision 12-May-1995 #text_change 21-Jul-2000
C: Accession: S51323; S51983; S71744
R: Bauer, A.; Koelling, R.
submitted to the EMBL Data Library, January 1995

RESULT 4

G02334
breast cancer tumor suppressor BRCA2 - human
N:Alternate names: breast cancer susceptibility protein BRCA2
C:Species: Homo sapiens (man)
C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 22-Jun-1999
C:Accession: G02334; S68501
R:Tavtigian, S.V.; Rommens, J.M.; Couch, F.J.; Neuhausen, S.; Bell, R.; Berry, S.; Bozde-
er, M.; Snyder, S.; Stringfellow, M.; Stroup, C.; Swedlund, B.; Teng, D.; Thomas, A.; Th
submitted to the EMBL Data Library, December 1995
A:Reference number: H01078
A:Accession: G02334
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-3418 <TAV>
A:Cross-references: EMBL:U43746; NID:g1161383; PIDN:AAB07223.1; PID:g1161384
R:Wooster, R.; Bignell, G.; Lancaster, J.; Swift, S.; Seal, S.; Mangion, J.; Collins, N.;
ith, A.; Connor, F.; Arason, A.; Gudmundsson, J.; Fienec, D.; Kelsell, D.; Ford, D.; To
; Narod, S.; Lenoir, G.; Egilsson, V.; Barkadottir, R.B.; Easton, D.F.; Bentley, D.R.
Nature 378, 789-792, 1995
A:Authors: Futreal, P.A.; Ashworth, A.; Stratton, M.R.
A:Title: Identification of the breast cancer susceptibility gene BRCA2.
A:Reference number: S68501; MUID:96112016
A:Accession: S68501
A:Molecule type: mRNA
A:Residues: 282-371, 'N', 373-598, 'S', 600-1108, 'EQ', 1111-1119, 'D', 1121-2321, 'V', 2323-2386;
C:Genetics:
A:Gene: GDB:BRCA2
A:Cross-references: GDB:387848; OMIM:600185
A:Map position: 13q12.3-13q12.3
C:Superfamily: breast cancer tumor suppressor BRCA2
C:Keywords: polymorphism; tumor suppressor

Query Match 4.7%; Score 145; DB 1; Length 3418;

Best Local Similarity 20.9%; Pred. No. 0.81;

Matches 96; Conservative 67; Mismatches 158; Indels 138; Gaps 23;

QY 36 KEDVAMELERVGEDEOMMLKRSECNPLQEPIDASAOFCATAGTCRKSVPCCWFRVVK 95
DB 972 KSDISLNDIKPEKNNDYMNKAG-----LLGPIHSNHSFGSPTASKEIKLSEHNKK 1026
QY 96 QRLFGTAGRFDVYFISPOGLKFRSKSSLANYLHKNGTSL-KPEDFDFTVLSKRGIKSR 154
DB 1027 SKWFFK-----DIEEQYPTSL---ACVEIVNTLALONQKLLKSPQIN-TVSAHLOSSV 1077
QY 155 YKDCSMAALTSHL--QNQSNNSNNLRTSKCKKDVMPSPSSSELOESRGLSNFTSTHL 212
DB 1078 VSDCKNSHITPOMLFKQDFNSNHL-----TPSQKAEITE-----LSTIL 1118
QY 213 LLKDEGVDDVDFNFRKVRK-----PKGKVTLK----- 239
DB 1119 ----EESGSGOFETQRRKPSYILOKSTFEVPENQMTILKTTSECRDADLHVTMNPSTG 1174
QY 240 -----GIPKK-----TKGCRKSCGFGVQSDSKRESVCKNKADESEPAVKOSL 284
DB 1175 QVDSSKQFEGTVELKRFACLLKNDCKNSASGYLTDE-----NEVGFGFYSAHTKL 1227
QY 285 D-RTVCISDAGAGETLSVTSENLSVKKKERSLSSGNSFCPSQKTSGLINKPCSAKDS 343
DB 1228 NVSTEALQKAVKLFSDIENISETS---AEVHPISLSSKCHD-----SVVSMF---K 1277
QY 344 HNEKYEDTFLESEE-----IGTKVEVVERKEHLHTDILKRGSEMDNN---CSF 388
DB 1278 HNDK---TVSEKNNKCOLILONNIEMTTGTVEEI-----TENYKRNTEEDNKNYTA 1327
QY 389 TRK-----DFTGEKIFQEDTIPRTOIRERRKTSYLFSSKY 423
DB 1328 SRNSHNLFEQDSSKNDTV---CIHKDETLLFTDQHN 1363

RESULT 5

S27833

rhoptry-associated protein 1 precursor - malaria parasite (Plasmodium falciparum)
N:Alternate names: protective antigen
C:Species: Plasmodium falciparum
C>Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Jun-2000
C:Accession: A45514; S27833
R:Ridley, R.G.; Takacs, B.; Lahm, H.W.; Delves, C.J.; Goman, M.; Certa, U.; Matile, H
Mol. Biochem. Parasitol. 41, 125-134, 1990
A:Title: Characterisation and sequence of a protective rhoptry antigen from Plasmodium
A:Reference number: A45514; MUID:90348711
A:Accession: A45514
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-782 <RI2>
A:Cross-references: GB:M32853; NID:g160656; PID:g160657
C:Superfamily: Plasmodium falciparum rhoptry-associated protein 1

Query Match 4.6%; Score 141.5; DB 2; Length 782;
Best Local Similarity 21.1%; Pred. No. 0.18;
Matches 110; Conservative 71; Mismatches 176; Indels 165; Gaps 25;

QY 123 SLANVLHKNGET---SLKPEDFDF-----TVLSKRGIKSRKDCSMAALTSHLQNS- 171
DB 18 NVADGINVNGDNNYCKTIINDFNFDYNYWTPINKKEFLNSYED--EFSSESFLNKSS 75
QY 172 -NNSWNLRTSKCKKDVMPSPSSSELOESRGLSNFTSTHLLKDEGVDDVDFNFRKVRK 230
DB 76 VDDGNINLTDSTSNK-----SSKKGHRSVRSASAAAILLEDDSKDDMEF----- 122
QY 231 PKGKVTILKGIPIKTKGCRKSCGFGVQSDSKRESVCKNKADESEPAVKQSOLDRTVCI 290
DB 123 -KASPSVVKVTSPTSGTQSLKSS---PSSTKSSPSN-----V 158
QY 291 SDAGACGETLSVTSEENSLVKKKERSLSSGNSFCSEQKTSGLINKFCSAKDSEHNKEDY 350
DB 159 KSASPHGES---NSSEESTTKSKRSAS-----VAGIV-----GADEEAPAPKN 200
QY 351 TPLESEETI-GTKVEVVERKEHLHT-----DILKRG-----EMDNNSPTRKDFGT 395
DB 201 TLTPLEELYPTNVNLFNYKYSLNMEENINILNKGDLVAQKEEFYDENMEKAKQDKKK 260
QY 396 --EKI-----FOEDTIPRTOIRERRKTSYLFSSKYNKEALSPPRKFAKFWTPPRS 443
DB 261 ALEKLGKSDSEDPFMFSENKFLNENOVKERNVAGSFRFSK--LNP-----FKK----- 307
QY 444 PENLYQETLFDHPWKLKLLIATIFLNRTSGKMAIPVLWKLEKYPASA----- 488
DB 308 -DEVTEKT---EVSKKTFSGIGFNLTDKKAKVLGVGATYQYETPMTLYNCNPNNLFDTI 363
QY 489 -----EVATADNRDVSSELLKPLGLYDLRAK-TIVKFSDEYLTQWKYPIEL 534
DB 364 ESLOGRIIDIKKRESMISTTFEQQKECLKNMGVLDELNDTQCKFGT----- 410
QY 535 HGIGKYGNDSYRIF-CVNEWKQVHP-----EDHKLNKYH 567
DB 411 -CIGSFGHEHLRLYEFDNDLKFHFHNIDYLTLDAGYKLOKNH 451

RESULT 6

S55785
nucleolar protein gar2 - fission yeast (Schizosaccharomyces pombe)
N:Alternate names: probable RNA-binding protein RBD18
C:Species: Schizosaccharomyces pombe
C>Date: 28-Oct-1996 #sequence_revision 07-Feb-1997 #text_change 17-Nov-2000
C:Accession: S55785; S68087; T37634; S52318
R:Gullis, M.P.; Girard, J.P.; Zabetakis, D.; Lapeyre, B.; Melese, T.; Caizergues-Perre
Nucleic Acids Res. 23, 1912-1918, 1995
A:Title: gar2 is a nucleolar protein from Schizosaccharomyces pombe required for 18S
A:Reference number: S55785; MUID:95319932
A:Accession: S55785
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-500 <GUL>

A:Cross-references: GB:222780; NID:g396104; PIDN:CAA80457.1; PID:g396105
C:Keywords: cytoskeleton

```

Query Match      4.5%; Score 136.5; DB 2; Length 598;
Best Local Similarity 23.3%; Pred. No. 0.27;
Matches 111; Conservative 72; Mismatches 205; Indels 89; Gaps 21;

QY  2 GTTGLESGLDRGAAPVTSSERLVPPPPNDLRKEDYAMELRVGEDEEQMMLKRSSEC 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  70 GTPPLKKUSKKGGYATNPESQIVBEKTRQNEAD-KTPLKSHNEQSKSKSSSET 128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  62 NPLOEPITASAOFGATAGTECRKSVPCGWVRVVKORLFGKTAGRDVYFI----SPOGLK 117
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  129 NP-----ESQNSKTVSKNSQ-----KDKDKSKSKKTNTFEHLTKNPKDL 171
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  118 FRSKSLANYLHKNGETSLKPDFTVLKRGKISRYKDCSMAALTSHLON--OSNNSN 175
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  172 KRSKTSNDPISICSENSL--NVDFLVLVGQ-----SDDESINFDWLRLNYSQNSKN 222
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  176 WNLRTSRCKCKDV--FMPSPSSSELOESR-GLSNFTSTHLLKDEGVDDV----- 223
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  223 YSLKYTKYTKDKTKNAKSDAESDSDAKKSKVKNVKKDKKDKKDKKDKKDKKDKK 282
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  224 -----NFRKVRKPKGVITLKGIPKTKK-----GCRKSCSGFYQS 260
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  283 ESGSKDKERKDPK-KDKKKLKKDKKDKTKYPESTDTESGDARNDNRNLKSKND 341
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  261 DSKRS--VCNKADAESEPAQKSLDRVTCISDAGAGETLSVTSENSLVKKERSLS 318
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  342 DKKDAKKTFTDSELESSESQDEKDKDKD-----SKTDNKKK-VKNDEESTD 392
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  319 SCSNFCSEKTSIINKFCSAKDSEHKEYEDT--FLESEIGTKVEVVERKEHLHTDIL 376
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  393 ADSEPKGDSK-KGKDEKGGKKDKKDKKDKKDKKDKKDKKDKKDKKDKKDKKDKK 448
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  377 KRGESENNCSPTKDKTGEKIFQEDTIPRTQIERRKTSYFSSKYNKEALSPRRK 433
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  449 KKGSKDKIK-KDARKD-TESTDAEDESCKTGF-KTSTKIKGSDTESLYKPKCAK 502
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
T34036
hypothetical protein B0041.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34036
R:Fulton, R.; Wohldmann, P.
submitted to the EMBL Data Library, April 1997
A:Description: The sequence of C. elegans cosmid B0041.
A:Reference number: Z21466
A:Accession: T34036
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1359 <FUL>
A:Cross-references: EMBL:AF000196; PIDN:AAC24256.1; GSPDB:GN00019; CESP:B0041.7
A:Experimental source: strain Bristol N2; clone B0041
C:Genetics:
A:Gene: CESP:B0041.7
A:Map position: 1
A:Introns: 12/2; 59/2; 248/2; 582/1; 675/1; 733/3; 966/2; 1044/2; 1310/3

Query Match      4.5%; Score 136.5; DB 2; Length 1359;
Best Local Similarity 19.2%; Pred. No. 0.82;
Matches 113; Conservative 91; Mismatches 253; Indels 131; Gaps 19;

QY  15 GAAPTVTSSERLVPPPPNDLRKEDYAMELRVGEDEEQMMLKRSSECNPLLEPTASQF 74
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  4 GVSESDSGCHVIED--EDL-----EMARQIENRERKRAQKLUKRE----- 44
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  75 GATAGTECRKSVPCGWVRVVKORLFGKTAGRDVYFISPOGLKFRSKSLANYLHKNGET 134
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db  45 -----RECKPPPKRPAKRRKASSEDDDDDEESPR--KSSKKSRKRAKSESSE 94
QY  135 SLKPEDFDFTVLKRGKISRYKDCSMAALTSHLQNSNNSNWLRTSRCKCKDVFEMPPSS 194
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  95 SDEEDKKKSKKKKKVQDKKKKKRKRITSSSEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 154
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  195 SSELQESRGLSNFTSTHLLKDEGVDDVNFVRKVRK-PKGKVTILK-----GI 241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  155 SSESSEEE-----RKVKSKKNKESKVKKRAFTSESDEDEK 191
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  242 PIKTKKGCGRKSKSGFVQSDSKRESVCNKADAESEPAQKSLDRVTCISDAGAGETLS 301
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  192 PSKSKKGLKAKKAKSESESESESESESESESESESESESESESESESESESESESESE 244
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  302 VTSENSLVKKKRSLSGSGNFCSEQKTSG-----IINKFCSAKDSEHKEYEDTF 352
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  245 TEKKRKRTSSESESESESESESESESESESESESESESESESESESESESESESE 298
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  353 LESREIGTKVEVVERKEHLHTDILKRGSEMDNCSPTKDKTGEKIFQ--EDTIPRTQIE 410
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  299 -----VEVLPQKK-----KRCVTLISDSEDEKDKSESEASDVEEKVKKKAK 342
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  411 RRKTSYFSSKYNKEALSPRRKAFKKTWPPRPFNLVQETLFDHPWKLLIATIFLNRTS 470
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  343 KQESS--ESGSDSSEGSITVNRKSKKKKPKKKKKGIIM-----DSSKLQKETIDAERA 395
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  471 GKMAIPVLWKFLE-KYPSAEVARTADWRDVSSELLKPLGLYDLRAKTIVKFSDEVLTQWK 529
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  396 KERR-----KRLEKKQKFNGLVEEGEDLTEMJTGTSORRKLKSVVLDPSSTVDEESK 450
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  530 YPFLHGTGKGYNDYRIFCVNEMKQVHPEDHKLKLYHDLWLNHEKL 577
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  451 KPVEVH-----NSLVRL-----KPH-QAHGIQPMYDCAFESLDRL 485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
T41622
probable ABC transporter - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T41622
R:Aert, R.; Volckaert, G.; Mcdougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, October 1999
A:Reference number: Z21735
A:Accession: T41622
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-822 <AER>
A:Cross-references: EMBL:AL122011; PIDN:CA858409.1; GSPDB:GN00068; SPDB:SPCC825.01
A:Experimental source: strain 972h-; cosmid c825
C:Genetics:
A:Gene: SPDB:SPCC825.01
A:Map position: 3

Query Match      4.4%; Score 135; DB 2; Length 822;
Best Local Similarity 22.4%; Pred. No. 0.52;
Matches 80; Conservative 53; Mismatches 134; Indels 90; Gaps 14;

QY  149 RGKIS--RYKINCSMAALTSHLQNSNNSNWLRTSRCKCKDVFEMPPSSSELOESRGLSN 206
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  3 RGRKSTQDADLELESLSQSESESPQP---VTKSKAKK-----NKKLNKASAFNS 51
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  207 FTSTHLLLK-DEGVDDVNFVRKVRKPKGVITLKGIPDKTKKGCGRKSKSGFVQSDSKRE 265
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  52 DNDSNYDLKPKDEDEVJ-----EVVPVKKPKSKSKAKANAFADFADEQSVEE 100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  266 SVCNKADAE-----SEPAQKSLDRVTCISDAGAGETLSVTSENSLVKKKERS 316
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  101 E--EEDSEKPVNRKKNKSKSKKASPKNAFDALDDMDLSDLEDESESESESESESESESE 158
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  317 LSSGNSNFCSEKTSIINKFCSAKDSEHKEYEDTFLESEIGTKVEVVERKEHLHTDIL 372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db   159 KDDG-----SEALDGDGI-----ESSEKDKKKKKXENDDAPKKDRKTRRKEEKAKRLA 208

Qy   373 TDILKRSEMNNNGSPTRKD-----FTGEKIFQEDTIPRTQTERRKTSL 416
      : | | | | | : | | | | | : | | | | | : | | | | |
Db   209 SMLSENKNDANAAPLNKTDAFKDGLPSGRLLFAYASGQKVADPGSNPADGI----- 261

Qy   417 YFSKKYNKEALSPPRRKAFKKWTPRSPFNLVQTLFDHPW-KULLIATIFI LNRTSGK 472
      |||| : | | | | | : | | | | | : | | | | | : | | | | |
Db   262 ---TVTGNTLLSPNSR-----DLQVEKLSVSAWGKLIIKDSELNLINGR 302

RESULT 11
A72682
C: probable A/G-specific adenine glycosylase APE0875 - Aeropyrum pernix (strain Kl)
C: Species: Aeropyrum pernix
C: Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C: Accession: A72682
R: Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahara, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Takahara, D. Nucleic Acids Res. 24, 63-101, 1996
A: Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix strain APE0875
A: Reference number: A72450; PMID:99310339
A: Accession: A72682
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-286 <RAW>
A: Cross-references: DDBJ:AP0000060; NTD:g5104188; PIDN:BAA79857.1; PID:d1043643; PID:g5104188
A: Experimental source: strain Kl
C: Genetics:
A: Gene: APE0875
```

	Query Match	Best Local Similarity	Score	DB 2;	Length	286;
Matches	28;	Conservative	26;	Mismatches	31;	Indels
						9; Gaps
						3;
QY	455	DPWKLLIATFLNRTSKMAIPVLWKEFKYPSEAVARTADWRDVSLEKPLGLGYDLRAK	514	: : :	: : :	:
Dδ	98	DPMAILVAAPFLLRKTTRQVRVVEEFLRRYPNFKAALASAREDEVRELIRPLGIEHORAK	157	: : :	:	: :
QY	515	TIVKFSDYEITKW--KYPI-----ELHGIGKY	540	::: :: : :	:	:
Dδ	158	HLTIELA-KHEARYGGRIPCSKKKLKLPGVDGI	190	::: :: : :	:	:

RESULT 12
A71928
cag island protein - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C:Accession: A71928
R:Alm, R.A.; Gibson, R.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:reference number: A71800; MUID:99120557
A:Accession: A71928
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1819 <ARN>
A:Cross-references: GB:AE001481; GB:AE001439; NID:g4155005; PIDN:AAD06047.1; PID:g415500
A:Experimental source: strain J99
C:Genetics:
A:Gene: orf13/14

```

Query Match          4.3%; Score 132; DB 2; Length 1819;
Best Local Similarity 18.2%; Pred. No. 2.4;
Matches 87; Conservative 83; Mismatches 183; Indels 124; Gaps 19;

QY 6 LESLSGDRCAAPTVTSSERLVPDPNDRKEDVAMELERVGED----- 49
    ||: ||

```

```

942 LENOALDCLNNAKTEABEKKRCVKDLPKDLQKVLAKEGSVRVYDCVSKAKNEABKECEK 1001
Qy 50 ---EEOGMMIKRSECNPLQIEPTIASQFATAGTECKRSVPCGWERVVKORLFGKTAGR 105
Db 1002 LLTPEARKLLEAEKESVKAYKDCVSRAR-NEKEKQECEKLLTPEARKLLEQEVKKSVKAY 1060
Qy 106 FDVY-----FISPOGLKFRSSSLANYLHKNGETSLK-----PEDPDF 143
Db 1061 LDCVSRAFNEKEKECEKLLTPEARKLLENQALD--CLNNAKTEABEKKRCVKDLPKDLQK 1118
Qy 144 TVLSKRGIKSRYKDCSMAALTSHLQONSNNSNNWLRTSRCKKDDVMPSPSSSFLQESRG 203
Db 1119 KVLAKESVKA-YLDC-----VSRARNE-----KKEKCEK--LLTPEARKLLEESK- 1161
Qy 204 LSNFTSTHL--LLKEDGVDVDFNRKVRKPKGVTKILGPIRKTKRG----- 250
Db 1162 --KSVKAYLDCVSKAKNEABKEKEKLLTPEARKLLEAEKESVKAYKDCVSRARNEKEKQ 1219
Qy 251 -----RKSCSFVOSDSKRESVCNKADAES--EPVA-----QKSOL 284
Db 1220 ECKLLTPEARKLLEQEVKKSVKAYLDCVSRARNEKEKECEKLLTPEARKFLFKORQOK 1279
Qy 285 DRTV--CI-----SDAGACGETLSVTSEENSVLKKBSRSLSSGNSFCSEQKTSINKFC 337
Db 1280 DKAIKDCLKNADPNDRAAIMKCLDGLSDEEKI-----KYLOEABERAVLDCLK 1327
Qy 338 SAKDSBHKEYYDFLE--SEEIGTKVEVWERKEHLHTDILKRGSE--MDNNSCSPTRKD 392
Db 1328 TARTDEEKRKCONLYSDLIOENKKAONKONOLSKTERLHOASECDLNDLDDPTDOE 1384

```

RESUI "T" 13
 T43523
 cut17 protein - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 21-Jan-2000 #sequence, revision 21-Jan-2000 #text_change 02-Jun-2000
 C:Accession: T43523; T41649; T41700
 R:Morishita, J.; Matsusaka, T.; Yanagida, M.
 submitted to the EMBL data Library, August 1999
 A:Description: Fission yeast cut17 is required for chromosome segregation.
 A:Reference number: Z22536
 A:Accession: T43523
 A:Status: translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 1-997 <MOR>
 A:Cross-references: EMBL:AB031034; PIDN:BA83415.1
 R:Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
 submitted to the EMBL data Library, August 1998
 A:Reference number: Z2007
 A:Accession: T41649
 A:Status: translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-997 <HAR>
 A:Cross-references: EMBL:AL031323; PIDN:CAA20434.1; GSPDB:GNO0068; SPDB:SPCC962.02c
 A:Experimental source: strain 972h-; cosmid c962
 R:Wedler, H.; Duesterhoeft, A.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
 submitted to the EMBL data Library, October 1999
 A:Reference number: Z22010
 A:Accession: T41700
 A:Status: translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 932-997 <WED>
 A:Cross-references: EMBL:AL121859; PIDN:CA58376.1; GSPDB:GNO0068; SPDB:SPCP31B10.10c
 C:Genetics:
 A:Gene: cut17; SPCC962.02c; SPDB:SPCP31B10.10c
 A:Map position: 3L
 A:Intons: 43/3

Query Match	4.3%	Score	131.5	DB 2:	Length	997
Best Local Similarity	20.1%	Pred. No.	1.1			
Matches	129	Conservative	119	Mismatches	241	
				Indels	153	
				Gaps	33	

Db 531 SRSMKKYQSEDEDEISEQKPRAKDKAASGOQSVGSISSRSOARKLQPNKNMSSSITPR 590
Qy 227 ---KVRKPKGKVTILKGIPIKTKKGRKSCSGFVQSDSKRESVCNKADAESEPVAKSQ 283
Db 591 SAASVPRPSGKVS-----NTSSGRRS-----DKSLAQSVPNFSELIKENTKPSSL 636
Qy 284 LDRTVCISDAGACGETLSVTSEENSLVKKE--RSLSSG-----SNFCSEOKTSGIIN 334
Db 637 AVKTTMRSOVKSSGRTKNI--KEDTLQRPRLRKSSSGNIDFTELSTLCSDDM---MVS 691
Qy 335 KFCSAKDSE--HNEKYEDTFLESEBI-GTKVEVVERKEHLHTDILKRGSEM----- 382
Db 692 LRVDSDISETLRNEEYDEPEAEPEEVLENAREEEVEELETILVFEDGNPMLSEAYEKVD 751
Qy 383 ---DNNGSPTRKDFTEKIFQEDTIPRT 407
Db 752 HSGENCSE-----FLPATVPTT 768

Search completed: January 15, 2002, 13:52:37
Job time: 406 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: January 15, 2002, 13:51:13 ; Search time 57.72 Seconds
(without alignments)
123.589 Million cell updates/sec

Title: US-09-724-296-37

Perfect score: 1662

Sequence: 1 MIFRFGVSNAMSLWDASPA.....ELSSIRGVRKGGALQWKS 317

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*

1: /cgn2.6/ptodata/2/iaa/5A_COMB.pep:*

2: /cgn2.6/ptodata/2/iaa/5B_COMB.pep:*

3: /cgn2.6/ptodata/2/iaa/6A_COMB.pep:*

4: /cgn2.6/ptodata/2/iaa/6B_COMB.pep:*

5: /cgn2.6/ptodata/2/iaa/PCTUS_COMB.pep:*

6: /cgn2.6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	93.5	5.6	1041	3	US-08-898-978-2
2	91.5	5.5	522	1	US-08-625-322-2
3	84.5	5.1	1073	4	US-09-541-782-6
4	84.5	5.1	1255	3	US-08-947-823-3
5	83.5	5.0	438	3	US-08-886-886-11
6	82.5	5.0	500	1	US-08-704-398-2
7	82.5	5.0	500	5	PCT-US95-05966-2
8	82.5	5.0	2710	1	US-08-480-604A-6
9	82.5	5.0	2710	2	US-08-405-496A-6
10	82.5	5.0	2710	4	US-08-915-136-6
11	82	4.9	323	6	5185254-4
12	81.5	4.9	892	1	US-07-977-434-12
13	81.5	4.9	892	1	US-08-458-819-12
14	81.5	4.9	892	5	PCT-US91-07035-12
15	79.5	4.8	859	1	US-07-854-596B-35
16	79	4.8	376	4	US-09-041-718-2
17	79	4.8	3224	2	US-08-705-660-34
18	79	4.8	3224	3	US-08-989-045-34
19	78.5	4.7	410	4	US-09-258-754-450
20	78.5	4.7	615	1	US-08-484-105-6
21	78.5	4.7	615	1	US-08-484-106-6
22	77.5	4.7	483	3	US-09-027-166-7
23	77	4.6	704	1	US-08-188-582-18
24	77	4.6	704	1	US-08-646-715-18
25	77	4.6	2293	4	US-09-368-590-2
26	76.5	4.6	772	1	US-08-524-757-12
27	75.5	4.5	396	6	5349058-2

28	75	4.5	704	1	US-08-188-582-5
29	75	4.5	704	1	US-08-646-715-5
30	75	4.5	1257	3	US-08-947-823-5
31	74.5	4.5	369	1	US-08-844-055-2
32	74.5	4.5	369	3	US-09-006-849-2
33	74.5	4.5	381	1	US-07-936-163-3
34	74.5	4.5	429	1	US-08-906-744A-2
35	74.5	4.5	429	3	US-09-093-134-2
36	74.5	4.5	541	3	US-08-890-615-2
37	74.5	4.5	976	4	US-09-104-324B-4
38	74.5	4.5	2522	4	US-09-251-645-13
39	74	4.5	263	4	US-09-159-106-2
40	74	4.5	303	4	US-09-159-106-13
41	74	4.5	435	4	US-09-159-106-11
42	74	4.5	436	4	US-09-150-213-4
43	74	4.5	2496	4	US-09-125-028-2
44	74	4.5	2958	4	US-08-894-344C-2
45	73.5	4.4	264	2	US-08-402-804-4

ALIGNMENTS

RESULT 1

US-08-898-978-2
; Sequence 2, Application US/08898978
; Patent No. 6001602
; GENERAL INFORMATION:
; APPLICANT: Reichard, Raymond W.
; APPLICANT: Brown, James R.
; APPLICANT: Lawlor, Elizabeth J.
; TITLE OF INVENTION: No. 6001602el iles
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/898,978
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: GM10051
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-898-978-2

Query Match 5.6%; Score 93.5; DB 3; Length 1041;
Best Local Similarity 19.7%; Pred. No. 0.39;
Matches 57; Conservative 42; Mismatches 104; Indels 87; Gaps 13;

QY	71	SIYPLATHPDVNW-----DFT-----PFOEFREJGELVKTHQLSTSPH	111
Db	234	SNMALVHPELTYVRIKDKESGDEYILGOESLPRFPFDRESYEWIQLSGKSLVGQSPY	293
QY	112	NOFTLTPSKESVTKNAVTDMAHYHMLRPMGAIADRSVIN-----THIGAYGNKDTA	164
Db	294	LFPYFDKKE-----LEAFRLPADFLEESEGTGIYHMAPAFCEADFF	336
QY	165	TQAFH-----ONIKOLPOETIKERMTLENDKTY---TTEETL	198
Db	337	ACQEHNVPLVCPDNOGCYTAEVKDFVGEYIKSAOKGIARR--LKNENKLFYQGTVRHRY	394
QY	199	QVCEQEDVPVFDEHHFYANPDHDNALVPRMKTWERIGLOPKVHLSSPKSEQARS	258
Db	395	PCWRTDLSLIYKAVNSW-----FVAVEKVKSKMLKANESIHWTEP-HIKQRPFGKWLEG	448
QY	259	HADYVDANELFERFROGWNIDFMIEAKOKAKALLRL--MDELSIRGVK	306
Db	449	ARDWA-----ISRNRVWGTPPIW---RSDGELAVIGISQIEALSGOK	490

```

RESULT      2
US-08-625-322-2
; Sequence 2, Application US/08625322
; Patent No. 5804412
; GENERAL INFORMATION:
; APPLICANT: Gill, Gordon N.
; APPLICANT: Kurten, Richard C.
; APPLICANT: Cadena, Deborah L.
; TITLE OF INVENTION: Sorting Nexins and Methods of Using Same
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/625,322
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-UD 1955
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 522 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-625-322-2

```

```

251 NHPMTLQD---PVRREFLEKEELPRAVGQTQLSGAGLLKMFNKATDAVSKMTIKMNESDI 307
QY 133 AYHYRMLEAMGIADR-----SVINIHHIGAYGNKDTA--TAQFHQNIKOLPOEIKERM 185
Db 308 WFEELQVECEBQRLKRLHAVVETLVN--HRKELALNTAQFAKSLAMLG----- 355
QY 186 ENDDKTYTTEETLQVCEQEDVPVDFHFHYANPD-----DH-----ADLNVALPRM 232
Db 356 SSEDNTALSRALSQLAEEVEEK--IEOLHGOEAOANPDFLLAELLSDYIRLLATVRAAFDQR 413
QY 233 IKTWERIGLPKVHLSSPKSEGAIRSHADYVDANFTLLERFRO-----WGTNIDEMIEAK 286
Db 414 MKTWQR-----WQDAQATLQKKREAREALLWANKPKLQQAQ 450
QY 287 QDKALLRLMDEISSIRGVKRIGGGALOWKS 317
Db 451 -----DEI-----LEWES 458

RESULT 3
US-09-541-782-6
; Sequence 6, Application US/09541782
; Patent No. 6284480
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/541,782
; CURRENT FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1073
; TYPE: PRT
; ORGANISM: S.pombe
US-09-541-782-6

```

Query Match	5.1%;	Score 84.5;	DB 4;	Length 1073;	
Best Local Similarity	19.9%;	Pred. No. 3.9;			
Matches	62;	Conservative 57;	Mismatches 130;	Indels 63;	Gaps
QY	28	YSKLSKTER- ---KEALLIVTKANLRNMTLHYII- ---GHGILPYRFSSSIIVPLA- T 77			
Db	571	YEKLDTERNNKENENNFWNLKFNLLTULRSFHGSEFTDTNIGYFTLLDNDFNASMEELNLT 630			
QY	78	HPDVMWDFVTPQKFERIGEIVKTHQLRTSPHPNQFTLFTSPKESVTKNAVDTMAYHYR 137			
Db	631	HSNOLLISMTKITEHFQSLDEALQSARSCCAPNSLDLIVSELKD- SKNSLLDALEH- - 687			
QY	138	MLEANGIADRSVINTHIGGAYGN- ----KDTATQAFHONTKOLPOETKERMWLENDDKT 191			
Db	688	-----SLQDISMSSOKLNGISSELEIQLKD- ----MKSESQOLVQELRSLYNLO- ---- 732			
QY	192	YTTEPTLQVCEQEDVPFVDFHHIFYANPDHADLNVALPRMKTIKTWERICLOPKVHLSSPK 251			
Db	733	HYHEES- ----QKEL- ----MYGVNRD- ----IDALVKTCTTSLUNDADIILSDYI 772			
QY	252	SEQATRSADHYDV- ----ANFLERFROWGTNIDFMIEAKOKDKALLRLMLDELSSI 302			
Db	773	SDOKSKFESKQODLIANTICKIYVSNFLOEONESLYTKADILHSHLNDTNSNTPKANEIWN 832			
QY	303	RGVKRIGGGALQ 314			
Db	833	RSEEFRLNAAO 844			

RESULT 4
US-08-947-823-3
; Sequence 3, Application US/08947823
; Patent No. 6114605


```

RESULT 6
US-08-704-398-2
; Sequence 2, Application US/08704398
; Patent No. 5679525
; GENERAL INFORMATION:
; APPLICANT: Peterson, Michael G
; APPLICANT: Henkel, Thomas
; TITLE OF INVENTION: EPSTEIN-BARR VIRUS TRANSCRIPTION FACTOR
; TITLE OF INVENTION: BINDING ASSAY
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOUBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,398
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/246,977
; FILING DATE: 20-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-592333/RAO
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-704-398-2

```

```

Query Match 5.0%; Score 82.5; DB 1; Length 500;
Best Local Similarity 21.6%; Pred. No. 1.9;
Matches 51; Conservative 32; Mismatches 68; Indels 85; Gaps 12;

QY 120 PKESVTKNAVTDMAHYRMLEAMGADRSVINIHGGA---YGN----- 160
Db 27 PKRLTREAMRN-----YKERG--DQVLLIHAKVAQKSYGNEKRFCCPPCVYLMGS 78

QY 161 -----KDTATAQFHQ-----NIKQLPQEIKERMTLENDKTYTTEETLQVCEQED 205
Db 79 GWKKKKQEOMERDGCSEQSPQCAFIGNSDQEQO---LNLEGNKYCTAKTLYISDSK 135

QY 206 -VPFVDFHFFYANPDHADLNVALPRMIKTWIGLQPKVHLSSP-KSQAIRSHADYV 263
Db 136 RKHFIFSVKMFYGNSD---DIGVFLSKRIK-----ISKPSKKKOSLK----- 175

QY 264 DANFLLERFRQWNTIDFMIEAKOKDKALLRLMDELSSIRGVKRIKG---GALOW 315
Db 176 -----NADLCIASGTVKVALFNRLRSQTVSTRYLHVEGNGFIASSQQW 217

```

```

RESULT 7
PCT-US95-05966-2
; Sequence 2, Application PC/TUS9505966
; GENERAL INFORMATION:
; APPLICANT: TULARIK, INC.
; TITLE OF INVENTION: EPSTEIN-BARR VIRUS TRANSCRIPTION FACTOR

```

```

; TITLE OF INVENTION: BINDING ASSAY
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOUBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/05966
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/246,977
; FILING DATE: 20-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: FP-59233-PC/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-05966-2

```

```

Query Match 5.0%; Score 82.5; DB 5; Length 500;
Best Local Similarity 21.6%; Pred. No. 1.9;
Matches 51; Conservative 32; Mismatches 68; Indels 85; Gaps 12;

QY 120 PKESVTKNAVTDMAHYRMLEAMGADRSVINIHGGA---YGN----- 160
Db 27 PKRLTREAMRN-----YKERG--DQVLLIHAKVAQKSYGNEKRFCCPPCVYLMGS 78

QY 161 -----KDTATAQFHQ-----NIKQLPQEIKERMTLENDKTYTTEETLQVCEQED 205
Db 79 GWKKKKQEOMERDGCSEQSPQCAFIGNSDQEQO---LNLEGNKYCTAKTLYISDSK 135

QY 206 -VPFVDFHFFYANPDHADLNVALPRMIKTWIGLQPKVHLSSP-KSQAIRSHADYV 263
Db 136 RKHFIFSVKMFYGNSD---DIGVFLSKRIK-----ISKPSKKKOSLK----- 175

QY 264 DANFLLERFRQWNTIDFMIEAKOKDKALLRLMDELSSIRGVKRIKG---CALOW 315
Db 176 -----NADLCIASGTVKVALFNRLRSQTVSTRYLHVEGNGFIASSQQW 217

```

```

RESULT 8
US-08-480-604A-6
; Sequence 6, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:

```

[illegible]

```

QY 172 IKOLPQEIKERMTLE - NDDKTYTYTEETLQVCQE -----DVFFVDFPHF---Y 216
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 227 SL---RKINSHGDIIRANSLFTEQELLNYSOELLNRGNLAAASDVRLLALKNFGVY 283
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 217 AN---PDHADLNVAPRM---IKTWERIGLOPKVHLSPKSEQAIRSHADYVDANEL 268
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 284 LDVMDLPQIGSDLPKTSIRSSGLDRWEMKL-----EAMKYKKYIN-NYT 330
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 269 LERF----RQNGTNIDFMIEAKOKDALLRLMEUS 300
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 331 SNNFQKLDQOLKDNFKLIIESKSEKSEIFSKLENLN 366
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

```

```

Db 112 EYTKQADIN--AEYNIKLWVDEAFVLT LKKAIVESSTTEALQLLEEEETQNPFQDNM 166
QY 133 AYHVRMLEAMGIADR-----SVNIHIGGAYGNKGTATATAQFHQN 171
Db 169 KFYKKRMEF--IYDRQKRFIVNYKSOINKPTVPTIDDIKSHLVSEYNRDET VLESYRTN 226
QY 172 IKQLPOEIKERWTLNDRKQTTYTEETLOVCEOE-----DVPPVFDFHHF--Y 216
Db 227 SL---RKINSHGIDIRANSLFTFEQELLNIYSOELLNRGNLAAASDIVRLALKNFGVY 283
QY 217 AN----PDQHDNLVALPRM---IKTWERIGLQPKVHLSSPKSQSAQIRSHADYVDANFL 268
Db 284 LDVDMPLGPHSDLFKFTISRPSSIGLDLRWEMIKL-----EAIMKYKKYIN-NYT 330
QY 269 LERF----ROWGNTIDFMIEAKQKDKALLRLMDLS 300
Db 331 SENFDKLDQOLKDNFKLIITESKSEKSEIFSKLENLN 366

RESULT 11
5185254-4
; Patent NO. 5185254
; APPLICANT: LINNENBACH, ALBAN
; TITLE OF INVENTION: GENE FAMILY OF TUMOR-ASSOCIATED ANTIGENS
; NUMBER OF SEQUENCES: 5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/291,583
; FILING DATE: 29-DEC-1988
; SEQ ID NO:4
; LENGTH: 323
5185254-4

```



```

; FILING DATE: 22-AUG-1986
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 746,121
; FILING DATE: 15-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US90/07641
; FILING DATE: 21-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 585,471
; FILING DATE: 20-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 455,611
; FILING DATE: 22-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 609,157
; FILING DATE: 02-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 557,517
; FILING DATE: 24-JUL-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Luann Cserr
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: Case No. 5795762 8753
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2972
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 892 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-458-819-12

```

Query Match 4.9%; Score 81.5; DB 1; Length 892;

Best Local Similarity 17.1%; Pred. No. 6.2; Mismatches 116; Indels 109; Gaps 11;

```

; Matches 58; Conservative 57; Mismatches 116; Indels 109; Gaps 11;
;
; Qy 31 LSKTERKEALLTVTKANLRNMTLHYIIGHGIPLYRFSSSIPLATHPDVMMDFVTPFQ 90
; Db 14 LNPDEKRNLEELSLKLYGKMFDFLGNVPLFCGDFSYPLERAVEYSCDADVTY 473
; Qy 91 KEFREIGELVYTHQVTSFHPNQFTL-----116
; Db 474 RIFKRLGRKIYENEMKLFYIEIEMELNGVYFDEYKLSKKYQEKMDGK 533
; Qy 117 ---FTSPKSVTKNAVTDMAHYRMLEAMGIADRSVINIHHGAYGNKDTATQAFHONIK 173
; Db 534 EKVEIAGETNLSSTOVA--YILFEKLNAP-----YKKTATGKFSTNAE 578
; Qy 174 QLPQEIKE---RMTLENDKTYTTEETLQVCEQEDVPFVD-----FHH----- 214
; Db 579 VLEELSKHEHIAKLLE---YKYOKLAKSTYIDSIPLSNRKNRNVHTFHTGTGTSTG 633
; Qy 215 --FYANP-----DDHADNLVALPRMIKTW-----ERICLOPKVHLSPPKSEQAI 256
; Db 634 RLSSNPMLQNLPRSEEGEIKRAVPRQDMMWILGADYSQIELRVLAHVS--KDNELL 691
; Qy 257 RSHADYVDAN-----FLLERFRWGTNIDFMI 283
; Db 692 KAFKEDLDIHTAAKIFGVSEMFVDSQMRVRGKMNFAI 731

```

RESULT 14

PCT-US91-07035-12

; Sequence 12, Application PC/TUS9107035

; GENERAL INFORMATION:

; APPLICANT: Gelfand, David H.

; APPLICANT: Abramson, Richard D.

; TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF

; TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES

; NUMBER OF SEQUENCES: 38

; CORRESPONDENCE ADDRESS:

```

; ADDRESS: Cetus Corporation
; STREET: 1400 Fifty-third Street
; CITY: Emeryville
; STATE: California
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/07035
; FILING DATE: 19910930
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,490
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,466
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,213
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 523,394
; FILING DATE: 15-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 143,441
; FILING DATE: 12-JAN-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 063,509
; FILING DATE: 17-JUN-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 899,241
; FILING DATE: 22-AUG-1986
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 746,121
; FILING DATE: 15-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US90/07641
; FILING DATE: 21-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 585,471
; FILING DATE: 20-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 455,611
; FILING DATE: 22-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 609,157
; FILING DATE: 02-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 557,517
; FILING DATE: 24-JUL-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Sias Ph.D. Stacey R.
; REGISTRATION NUMBER: 32,630
; REFERENCE/DOCKET NUMBER: Case No. 2580
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-420-3300
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 892 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US91-07035-12

```

Query Match 4.9%; Score 81.5; DB 5; Length 892;

Best Local Similarity 17.1%; Pred. No. 6.2; Mismatches 116; Indels 109; Gaps 11;

Matches 58; Conservative 57; Mismatches 116; Indels 109; Gaps 11;

Qy 31 LSKTERKEALLTVTKANLRNMTLHYIIGHGIPLYRFSSSIPLATHPDVMMDFVTPFQ 90

GenCore version 4.5
Copyright (C) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:50:07 ; Search time 120.01 Seconds
(without alignments)
195.661 Million cell updates/sec

Title: US-09-724-296-37

Perfect score: 1662

Sequence: 1 MIFRFGFVSNAWLDASPA.....ELSSIRGVKRGCGALQWKS 317

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_1101.*

1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	359.5	21.6	371	21 AAY44499	S. pombe delta228-
2	359.5	21.6	600	21 AAY44500	GST signal peptide
3	359.5	21.6	828	21 AAY44498	GST signal peptide
4	345	20.8	656	17 AA001618	Neurospora crassa
5	313.5	18.9	626	21 AAY44502	N. crassa delta228
6	313.5	18.9	626	21 AAY44503	B. subtilis delta22
7	292.5	17.6	294	21 AAY44505	D. radiodurans del
8	103	6.2	964	18 AA032619	Cyclic-isomaltolig
9	93.5	5.6	1040	20 AA097216	An isoleucyl tRNA
10	93.5	5.6	1041	21 AAY56782	C. trachomatis ile
11	92.5	5.6	234	21 AAB03759	Human H37 amino ac

12	92.5	5.6	674	21 AAB42482	Human ORFX ORF2246
13	92.5	5.6	674	21 AAB03758	Human H37 amino ac
14	91.5	5.5	522	19 AAY77094	Sorting nexin 1.
15	91	5.5	558	21 AAY77979	A. thaliana enviro
16	91	5.5	1284	9 AAP81187	Sequence encoded b
17	90.5	5.4	652	19 AAW73029	Helicobacter pylor
18	89	5.4	261	18 AAW55332	H. pylori ORF hp3e
19	88.5	5.3	537	21 AAG14498	Arabidopsis thalia
20	88.5	5.3	549	21 AAG14497	Arabidopsis thalia
21	88	5.3	289	22 AAG81570	S. epidermidis ope
22	88	5.3	645	19 AAW73030	Helicobacter pylor
23	87.5	5.3	548	20 AAY37206	Protein involved i
24	87	5.2	1521	21 AAG39235	Arabidopsis thalia
25	87	5.2	1528	21 AAG39234	Arabidopsis thalia
26	87	5.2	1562	21 AAG39233	Arabidopsis thalia
27	87	5.2	1703	21 AAG36714	Arabidopsis thalia
28	87	5.2	1710	21 AAG36713	Arabidopsis thalia
29	87	5.2	1744	21 AAG36712	Arabidopsis thalia
30	85	5.1	398	22 AAG81475	S. epidermidis ope
31	85	5.1	398	22 AAG82202	S. epidermidis ope
32	84.5	5.1	249	18 AAW55445	H. pylori ORF Olae
33	84.5	5.1	249	20 AAY17212	H. pylori outer me
34	84.5	5.1	458	21 AAY97013	S. cerevisiae esse
35	84.5	5.1	1255	19 AAW55974	Tomato pest resist
36	83.5	5.0	438	22 AAG81537	S. epidermidis ope
37	83.5	5.0	1254	11 AAR07503	Merozoite apical-e
38	83.5	5.0	1254	18 AAW24575	Merozoite apical-e
39	83	5.0	806	22 AAB88467	Human membrane or
40	83	5.0	1619	22 AAU00970	Drosophila melanog
41	83	5.0	1619	22 AAU00986	Drosophila melanog
42	83	5.0	1619	22 AAU00987	Drosophila melanog
43	83	5.0	1619	22 AAU00988	Drosophila melanog
44	83	5.0	1704	22 AAU00969	Drosophila melanog
45	83	5.0	1704	22 AAU00983	Drosophila melanog

ALIGNMENTS

RESULT 1
AAY44499
ID AAY44499 standard; Protein; 371 AA.
XX
AC AAY444499;
XX
DT 27-MAR-2000 (first entry)
XX
DE S. pombe delta228-UV damage endonuclease.
XX
KW Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;
KW glutathione-S-transferase signal peptide; uvel+ gene product;
KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
KW abasic site; platinum diaduct; mismatched nucleotide pairing;
KW nucleotide alkylation; skin cancer.
XX
OS Schizosaccharomyces pombe.
XX
PN WO9963828-Al.
XX
PD 16-DEC-1999.
XX
PF 08-JUN-1999; 99WO-US12910.
XX
PR 08-JUN-1998; 98US-0088521.
XX
PR 18-MAY-1999; 99US-0134752.
PA (UYEM-) UNIV EMORY.
XX
PI Doetsch PW, Kaur B, Avery AM;
XX WPI; 2000-116417/10.
DR N-PSDB; AAZ29859.
XX

PT A new truncated ultraviolet damage endonuclease for treatment of skin
 XX cancers -
 PS Claim 13; Page 53; 133pp; English.
 XX
 CC The present sequence is delta228-UV damage endonuclease. Delta228-UVDE is
 CC a 288 amino acid deletion of the N-terminal of the S. pombe uvel+ gene
 CC product. This is expressed in frame with a GST leader sequence to
 CC generate a fusion protein. This provides stable endonuclease fragments
 CC for cleaving a double-stranded DNA molecule that has a distorted
 CC structure resulting from UV radiation damage, a photoproduct, an abasic
 CC site, mismatched nucleotide pairing, aplatium diaduct, an intercalated
 CC molecule or alkylation of a nucleotide. Uvelp can be used in compositions
 CC for internal or topical application and as a therapeutic agent for skin
 CC cancers.
 XX
 XX Sequence 371 AA;
 SQ
 Query Match 21.6%; Score 359.5; DB 21; Length 371;
 Best Local Similarity 29.7%; Pred. No. 2.6e-28;
 Matches 83; Conservative 60; Mismatches 111; Indels 25; Gaps 7;
 Qy 34 TERKEALLTVTKANLRNTM---RTLHYIIGHGIPLYRFSSSIPLATHPDVMDVFTPF- 89
 Db 48 tiqrdglesvkqlgtqnvldlikivewhnhfghfmrsvsdlfpfash--akygytlefa 105
 Qy 90 QKEFREIGELVKTHTLTSFHPNQFTLTSPKESVTKNVTDMAHYRMLEAMGIADR-- 147
 Db 106 qshleevgklankynhrhltmhpggytqiasprevvvdsairdlayhdeilsrmkineqln 165
 Qy 148 --SVINIHIGGAYGNKDTATAQFHNKIQLPQEIKERMTLENDDKTYTTEFLQVCEQED 205
 Db 166 kdavliihlggtfegkktldfrknyqrlsdsvkarivlendddvswsqdlplcqln 225
 Qy 206 VPEVDFDHFHYANP---DDHADLNVALPRMIKTWERIGLQPKVHLSSPKSQAI----- 256
 Db 226 iplvldwhhnhivpgtiresgldlmpitretwrtkgtkqkghysesadptaisgmkr 285
 Qy 257 RSHADYVDANFLLERFROWGTNIDFMIEAKOKKALLRL 295
 Db 286 rahsdrv-----fdfpccpdtmdlmieakekeqavfel 318
 RESULT 2
 AAY44500
 ID AAY44500 standard; Protein: 600 AA.
 XX
 AC AAY44500;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 XX GST signal peptide and delta228 S. pombe UVDE fusion protein.
 DE GST signal peptide; glutathione-S-transferase signal peptide;
 KW ultraviolet damage endonuclease; UVDE; delta228-UVDE;
 KW repair-deficient E. coli strain; UV irradiation; DNA damage;
 KW UV radiation damage; photoproduct; abasic site; aplatium diaduct;
 KW mismatched nucleotide pairing; nucleotide alkylation;
 KW fusion protein; skin cancer.
 XX Schizosaccharomyces pombe.
 OS Synthetic.
 XX
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 11 /note= "Encoded by aaaa"
 FT Misc-difference 85 /note= "Encoded by gtt"
 FT Misc-difference 147 /note= "Encoded by gtt"
 FT
 XX W09963828-A1.

XX
 PD 16-DEC-1999.
 XX
 PF 08-JUN-1999; 99WO-US12910.
 XX
 PR 08-JUN-1998; 98US-0088521.
 PR 18-MAY-1999; 99US-0134752.
 XX
 PA (UYEM-) UNIV EMORY.
 XX
 PI Doetsch PW, Kaur B, Avery AM;
 XX WPI; 2000-116417/10.
 DR N-PSDB; AAZ29860.
 XX
 PT A new truncated ultraviolet damage endonuclease for treatment of skin
 PT cancers -
 XX
 PS Disclosure; Page 56; 133pp; English.
 XX
 CC The present sequence is a fusion protein comprising the GST signal
 CC peptide and delta228 truncated ultraviolet damage endonuclease (UVDE)
 CC from S. pombe. Delta228-UVDE is a 288 amino acid deletion of the
 CC N-terminal of the S. pombe uvel+ gene product. This is expressed in frame
 CC with a GST leader sequence. Stable endonuclease fragments can be produced
 CC for cleaving a double-stranded DNA molecule that has a distorted
 CC structure resulting from UV radiation damage, a photoproduct, an abasic
 CC site, mismatched nucleotide pairing, aplatium diaduct, an intercalated
 CC molecule or alkylation of a nucleotide. Uvelp can be used in compositions
 CC for internal or topical application and as a therapeutic agent for skin
 CC cancers.
 XX
 XX Sequence 600 AA;
 SQ
 Query Match 21.6%; Score 359.5; DB 21; Length 600;
 Best Local Similarity 29.7%; Pred. No. 5.5e-28;
 Matches 83; Conservative 60; Mismatches 111; Indels 25; Gaps 7;
 Qy 34 TERKEALLTVTKANLRNTM---RTLHYIIGHGIPLYRFSSSIPLATHPDVMDVFTPF- 89
 Db 277 tiqrdglesvkqlgtqnvldlikivewhnhfghfmrsvsdlfpfash--akygytlefa 334
 Qy 90 QKEFREIGELVKTHTLTSFHPNQFTLTSPKESVTKNVTDMAHYRMLEAMGIADR-- 147
 Db 335 qshleevgklankynhrhltmhpggytqiasprevvvdsairdlayhdeilsrmkineqln 394
 Qy 148 --SVINIHIGGAYGNKDTATAQFHNKIQLPQEIKERMTLENDDKTYTTEFLQVCEQED 205
 Db 395 kdavliihlggtfegkktldfrknyqrlsdsvkarivlendddvswsqdlplcqln 454
 Qy 206 VPEVDFDHFHYANP---DDHADLNVALPRMIKTWERIGLQPKVHLSSPKSQAI----- 256
 Db 455 iplvldwhhnhivpgtiresgldlmpitretwrtkgtkqkghysesadptaisgmkr 514
 Qy 257 RSHADYVDANFLLERFROWGTNIDFMIEAKOKKALLRL 295
 Db 515 rahsdrv-----fdfpccpdtmdlmieakekeqavfel 547
 RESULT 3
 AAY44498
 ID AAY44498 standard; Protein: 828 AA.
 XX
 AC AAY44498;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE GST signal peptide and S. pombe UVDE fusion protein.
 KW GST signal peptide; ultraviolet damage endonuclease; UVDE; Uvelp;
 KW UV irradiation; DNA damage; UV radiation damage; fusion protein;
 KW skin cancer; glutathione-S-transferase.

[illegible]

```

Db 511 --mtlppcpdmdlmieakkeqavfelm 537
RESULT 5
ID AAY44502 standard; Protein; 626 AA.
XX AAY44502;
AC AAY44502;
XX 27-MAR-2000 (first entry)
XX N. crassa delta228-UV damage endonuclease.
XX Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;
KW glutathione-S-transferase signal peptide; uvel+ gene product;
KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
KW abasic site; platinum diaduct; mismatched nucleotide pairing;
KW nucleotide alkylation; skin cancer.
XX Neurospora crassa.
XX WO9963828-A1.
XX 16-DEC-1999.
XX 08-JUN-1999; 99WO-US12910.
XX 08-JUN-1998; 98US-0088521.
XX 18-MAY-1999; 99US-0134752.
XX (UYEM-) UNIV EMORY.
XX Doetsch PW, Kaur B, Avery AM;
XX WPI; 2000-116417/10.
XX A new truncated ultraviolet damage endonuclease for treatment of skin
XX cancers -
XX Claim 16; Page 59; 133pp; English.
XX The present sequence is delta228-UV damage endonuclease from N. crassa.
XX Delta228-UVDE is a 288 amino acid deletion of the N-terminal of the
XX uvel+ gene product. This is expressed in frame with a GST leader sequence
XX to generate a fusion protein. This provides stable endonuclease fragments
XX for cleaving a double-stranded DNA molecule that has a distorted
XX structure resulting from UV radiation damage, a photoproduct, an abasic
XX site, mismatched nucleotide pairing, platinum diaduct, an intercalated
XX molecule or alkylation of a nucleotide. Uvelp can be used in compositions
XX for internal or topical application and as a therapeutic agent for skin
XX cancers.
XX Sequence 626 AA;
XX Query Match 18.9%; Score 313.5; DB 21; Length 626;
XX Best Local Similarity 30.5%; Pred. No. 3.1e-23;
XX Matches 82; Conservative 49; Mismatches 97; Indels 41; Gaps 9;
QY 46 ANLRNMTRLHYLIIGHGIPLYRFSSSIVPLATHPDVMDVFTPFQKE-FREIGELVKTHQ 104
Db 277 anardivkmlcwf-----pfashp-vhgyklapfasevlaeagrvaalg 320
QY 105 LRTSFHPNQFTLTSPKESVTNAVTDMAHYRMLAEMGIADR-----SVINIHIGGAYGN 160
Db 321 hrltthpgqftlgsprkvesairdleydhellslklpeqqrndavmiimmgqgfd 380
QY 161 KDTATAQFHNKQLPQETKERMTLENDKTYTTEETLQVCEQEDVPFDFHHFYANPD 220
Db 381 kaatlerkrnyarisqsknrivlendvgwtvhdllpvcceelnpmvldyhhnncfd 440
QY 221 D-----HADLNVALP-----RWIKTWERIGLOPKVHLSSPKSECAI-----RSHADYVDANF 267

```

```

Db 441 pahlregtidisdplqeriantwkrkigikmhysep-cdgavtprdrkrhrprv----- 495
QY 268 LLERFPOWCTNIDFMLEAKQKDKALLRLM 296
Db 496 --mtlppcpdmdlmleakkeqavfelm 522
RESULT 6
ID AAY44503 standard; Protein; 626 AA.
XX AAY44503;
AC AAY44503;
XX 27-MAR-2000 (first entry)
XX B. subtilis delta228-UV damage endonuclease.
XX Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;
KW glutathione-S-transferase signal peptide; uvel+ gene product;
KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
KW abasic site; platinum diaduct; mismatched nucleotide pairing;
KW nucleotide alkylation; skin cancer.
XX Bacillus subtilis.
XX WO9963828-A1.
XX 16-DEC-1999.
XX 08-JUN-1999; 99WO-US12910.
XX 08-JUN-1998; 98US-0088521.
XX 18-MAY-1999; 99US-0134752.
XX (UYEM-) UNIV EMORY.
XX Doetsch PW, Kaur B, Avery AM;
XX WPI; 2000-116417/10.
XX A new truncated ultraviolet damage endonuclease for treatment of skin
XX cancers -
XX Claim 16; Page 59; 133pp; English.
XX The present sequence is delta228-UV damage endonuclease from B. subtilis.
XX Delta228-UVDE is a 288 amino acid deletion of the N-terminal of the
XX uvel+ gene product. This is expressed in frame with a GST leader sequence
XX to generate a fusion protein. This provides stable endonuclease fragments
XX for cleaving a double-stranded DNA molecule that has a distorted
XX structure resulting from UV radiation damage, a photoproduct, an abasic
XX site, mismatched nucleotide pairing, platinum diaduct, an intercalated
XX molecule or alkylation of a nucleotide. Uvelp can be used in compositions
XX for internal or topical application and as a therapeutic agent for skin
XX cancers.
XX Sequence 626 AA;
XX Query Match 18.9%; Score 313.5; DB 21; Length 626;
XX Best Local Similarity 30.5%; Pred. No. 3.1e-23;
XX Matches 82; Conservative 49; Mismatches 97; Indels 41; Gaps 9;
QY 46 ANLRNMTRLHYLIIGHGIPLYRFSSSIVPLATHPDVMDVFTPFQKE-FREIGELVKTHQ 104
Db 277 anardivkmlcwf-----pfashp-vhgyklapfasevlaeagrvaalg 320
QY 105 LRTSFHPNQFTLTSPKESVTNAVTDMAHYRMLAEMGIADR-----SVINIHIGGAYGN 160
Db 321 hrltthpgqftlgsprkvesairdleydhellslklpeqqrndavmiimmgqgfd 380
QY 161 KDTATAQFHNKQLPQETKERMTLENDKTYTTEETLQVCEQEDVPFDFHHFYANPD 220

```

Db	73	lagdtdgaavtlthlapqlleaghaftdagvrlhnhpeqfivlnsdprvessvramisah	132
QY	136	YRMLEAMGIDRSVINITHIGAYGNKDTATAQFHNLIKQLPOEIKERMTLENDKTVTTE	195
Db	133	arvmdg19la-rfpwnlll--lhggkggrgaeaalipdipdvrlrlglenderayspa	189
QY	196	ETLOVCQEDVPVFDPHFHYANP--DDHADLANVA--LPRMKTWERIGHQPK- ---VHL	247
Db	190	elipiccatgplvfdahhvhvhdiklpdqedpsvrewlrlaratw----qpewagvqvl	244

Qy	248	SS	---	PKSEQAIRSIADYVDANFLLEFRQWGTNIDFMIEAKQKD	KAL--LRLM	296
				: : : : : : : : : : : : : : : : : :		
Db	245	sngiegppqdrhrhliadfp	sa-----yadvpqieavakqkeeaiaa	rlm	290	

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

AAW32619
ID AAW32619 standard: Protein: 064 AA

XX
AC AAW32619.

XX	
DT	
XX	04-FEB-1998 (first entry)
XX	
DE	Cyclic-isomaltoligosaccharide synthase.
XX	
KW	Cyclic-isomaltoligosaccharide synthase; cycloisomaltoheptose;
KW	anti decay activity; Escherichia coli; E.coli; microorganism;
KW	recombinant DNA.

KW anti decay activity; *Escherichia coli*; *E. coli*; microorganism;
KW recombinant DNA

XX
OC
Pacific
XXXX
DN
TD00334072-AXX
00-SEP-1997XX
REF ID: A68888XX
01-MAR-1966.

XX
PA (NOPO) NORINCITSANCUO CUOVIMIN COCO

PA (NODA) ZH NODA SANGYO KAGAKU KENKYUSHO.
YY

DR WPI; 199/-49/314/46.
DR N-PSDB: AATQ2376

XX
PT Cyclic-iso:malt:oligosaccharide synthase - used in the production of
PT cyclo:iso:malto:heptose with strong anti-decay activity

pH
cyclo: iso: malto: heptose with strong

Claim 3: Page 8-10; 12pp; Japanese.

The present sequence represents a novel cyclic-isomaltotigosaccharide synthase isolated from a *Bacillus* sp. (possibly *Bacillus circulans* U-155). The cyclic-isomaltotigosaccharide synthase has the following physicochemical properties: (1) action: acts on glucose polymers (e.g. dextran) consisting of 1,6- α -linkages to form cyclic-isomaltotigosaccharides based on cycloisomaltoheptose; (2) substrate specificity: acts on dextran having 1,6- α -linkages as main chains, but not on amylopectin, pullulan; and (3) optimum pH and stable pH range: active around pH 6.0 and stable in the range of pH 4.5 to 9.0. The DNA sequence may be used to produce cyclic-isomaltotigosaccharide synthase in microorganisms of the genus *Escherichia*, particularly *E. coli*. The resultant cyclic-isomaltotigosaccharide, synthase can efficiently produce cycloisomaltoheptose which has very strong anti-decay activity.

The DNA sequence m

CC	efficiently produce	6.2%;	Score 103;	DB 18;	Length 964;
CC	cycloisomaltotriose which has very strong	22.1%;	Pred. No. 0.24;		
CC	anti-decay activity.				
XX					
SQ	Sequence 964 AA;				
	Query Match				
	Best Local Similarity				

```

Query Match          6.2%; Score 103; DB 18; Length 964;
Best Local Similarity 22.1%; Pred. No. 0.24;
Matches 55; Conservative 44; Mismatches 84; Indels 66; Gaps
QY 95 EICELVKTHQIIRTFHPNQIITFTSPKESVTKNAVTDMAHYHMYMLEANG--IADRSVIN- 151

```

Matches	55;	Conservative	44;	Mismatches	84;	Indels	66;	Gaps	14;
QY	95	EIGELVKTHOLRTSFHPNOETLFTSPKESVTKNAVTDMAYHYRMLEAMG	--	IADRSVIN	--	IADRSVIN	--	151	

```

Db      561 elge--dsqmlaheypnr-----skmrstlksamkd---hynfitayenallfdadvldn 611
Qy      152 -----THIGGAYGNKDTATQAQFQNIKOLPO-EIKERMTELENDKTY-----TTEET- 197
Db      612 dagkqfinlagvntspdgaaantvwmkskrtpetpynllhlinlvnndqwrnsngqtaqtn 671
Qy      198 ----LOVCEQEDVPFVDFHHFYANPDHDLNVALP-----RMIKTWE 237
Db      672 latkvlygaeetigv-----yaaspdnggatqslpfttgtdssgsyisftvpsleyws 726
Qy      238 RIGLOPKVHLSSP-----KSEQATRS-----HADVVDANFLERFRONGTNIIDFMIEA 285
Db      727 mlym--krstaapvdmnyeataiksnvntnhagytgsgf-vdqtatvndvgvsvflvha 783
Qy      286 KQDKKALLR 294
Db      784 sskddyvlr 792

RESULT 9
AAW97216
ID AAW97216 standard; Protein; 1040 AA.
XX AC AAW97216;
XX XX
XX 06-MAY-1999 (first entry)
XX DE
XX XX
XX An isoleucyl tRNA (ileS) synthetase.
KW Isoleucyl tRNA synthetase; ileS; vaccine; ocular; genital trachoma;
KW lymphogranuloma venerum; keratitis; corneal infiltration; prostaticitis;
KW bacterial adhesion; in-dwelling device; cell invasion.
XX OS
XX Chlamydia trachomatis.
XX FH
XX Key Location/Qualifiers
XX FT Misc-difference 1 /note= "encoded by GTG"
XX XX
XX EP893495-A2.
XX PN
XX XX
XX 27-JAN-1999.
XX PD
XX XX
XX 10-JUL-1998; 98EP-0305502.
XX PF
XX XX
XX 23-JUL-1997; 97US-0898798.
XX PR
XX XX
XX (SMIK ) SMITHKLINE BEECHAM CORP.
XX PA (SMIK ) SMITHKLINE BEECHAM PLC.
XX XX
XX Brown JR, Lawlor EJ, Reichard R;
XX PI
XX WPI; 1999-097780/09.
XX DR N-PSDB; AAX15503.
XX XX
XX New Chlamydia trachomatis isoleucyl tRNA synthetase polypeptide and
XX polynucleotide - useful as diagnostic reagents and for prevention
XX and treatment of Chlamydia infections which cause ocular trachoma
XX and infertility
XX PS
XX Claim 6; Page 9-10; 34pp; English.
XX XX
XX The present sequence represents an isoleucyl tRNA (ileS) synthetase
XX polypeptide. Iles polypeptides are useful for screening for specific
XX inhibitors or activators, in vaccines to induce an immunological
XX response, to raise antibodies and for treating conditions associated
XX with a lack of ileS protein. Iles antagonists are used to treat
XX conditions requiring inhibition of ileS, specifically infection
XX by C. trachomatis, including ocular or genital trachoma, lymphogranuloma
XX venerum, keratitis, corneal infiltration and prostaticitis etc. Iles
XX polypeptides, polynucleotides and their antagonists can also be used to
XX prevent adhesion of bacteria, especially Gram-positive, to extracellular

```

```

CC matrix proteins in wounds or on in-dwelling devices, and to block ileS
CC polypeptide-mediated cell invasion or normal progression of infection.
XX
XX SQ Sequence 1040 AA;

```

Query Match 5.6%; Score 93.5; DB 20; Length 1040;

Best Local Similarity 19.7%; Pred. No. 2.6; Mismatches 42; Indels 87; Gaps 13;

Matches 57; Conservative 42; Mismatches 42; Indels 87; Gaps 13;

Qy 71 STVPLATHPDVMMW-----DFVT-----PFQKEFREIGELVKTHQLRTSHP 111

Db 234 smmalavhellyvrikdkesdeylqgeslprwfpdresyewigqslgksalvqsyep 293

Qy 112 NQFTLTSPKESVTNAVTDMAHYHRYMLEAMCIADRSVIN-----THIGGAYGNKDTA 164

Db 294 -lfpyfqdkke-----leafrilpadfleeegtgivhmapafgeadff 336

Qy 165 TAQFH-----QNIKOLPQEIKERMTLENDKTY---TTEETL 198

Db 337 acqehnvplvepvdnggcytaevkdfvgeyiksadkgiarr--lknenklfyqgtvhrhy 394

Qy 199 QVCEQEDVPFVDFHHFYANPDHDLNVALPRMIKTWERIGLOPKVHLSSPKSEQAIRS 258

Db 395 pfcwrtdspilykavns-----fvavekkskmlkanesihwtpe-hikgrfgrgkwieg 448

Qy 259 HADYVDANFLERFRONGTNIIDFMIEAKOKKALLRL--MDELSSIRGVK 306

Db 449 ardwa-----isrnywgtpipiw-----rsddgellivigsiglealsgqk 490

RESULT 10

AAW56782

ID AAW56782 standard; Protein; 1041 AA.

XX AC AAW56782;

XX XX

XX 27-MAR-2000 (first entry)

XX DE

XX C. trachomatis ileS polypeptide.

XX KW Isoleucyl tRNA synthetase; ileS; antibiotic; antibacterial; vaccine;

XX KW ocular trachoma; lymphogranuloma venerum; inclusion conjunctivitis;

XX KW mucopurulent rhinitis; urethritis; tropical bubo; cervical follicle;

XX KW prostaticitis; proctitis; infant pneumonitis.

XX OS Chlamydia trachomatis.

XX XX

XX US6001602-A.

XX XX

XX 14-DEC-1999.

XX XX

XX 23-JUL-1997; 97US-0898978.

XX XX

XX 23-JUL-1997; 97US-0898978.

XX PR

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX PA

XX Brown JR, Lawlor EJ, Reichard RW;

XX PI

XX WPI; 2000-096135/08.

XX DR N-PSDB; AAZ46652.

XX XX

XX Nucleic acids encoding Chlamydia trachomatis isoleucyl tRNA synthetase

XX polypeptides useful for treating, for example, urethritis, proctitis

XX and prostaticitis -

XX PS Claim 1; Columns 9-11; 21pp; English.

XX XX

XX This represents a isoleucyl tRNA synthetase (ileS) polypeptide derived

XX from Chlamydia trachomatis. The ileS polypeptide can be expressed by

XX standard recombinant methodology. The protein produced may then be used

XX to screen for antibiotics active against C. trachomatis and to study the

XX CC

CC role of the iles protein in the pathogenesis of disease. They are
 CC particularly useful as antigens in the production of antibodies specific
 CC for C. trachomatis. The antibiotics and antibodies obtained are useful
 CC for treating a wide range of diseases including classic ocular trachoma,
 CC lymphogranuloma Venereum, inclusion conjunctivitis, mucopurulent
 CC rhinitis, urethritis, tropical bubo, cervical follicles, prostatitis,
 CC proctitis and infant pneumonitis. The antibodies may also be used to
 CC detect the presence of C. trachomatis and related microbes according to
 CC standard assay protocols.

XX Sequence 1041 AA;

Query Match 5.6%; Score 93.5; DB 21; Length 1041;
 Best Local Similarity 19.7%; Pred. No. 2.6;
 Matches 57; Conservative 42; Mismatches 104; Indels 87; Gaps 13;

Qy 71 SIVPLAPHDPMV-----DFVT-----PFOKEFREIGELVKTHQRTSFHP 111
 | : | | :
 Db 234 snmalavhpeltyvrikkdesgdeylilgqeslprwfpdresyewilglsksgivggsyep 293

Qy 112 NQFTLTSFKESVTKNAVDMATHYRMLBAMGADRSVIN-----IHGGAYGNKDTA 164
 | : | | :
 Db 294 -lfpyfgdkke-----leafriipadfielsegtgihmapafgeadff 336

Qy 165 TAQPH-----ONIKQLPOEIKERMTELENDKTY---TTEETL 198
 :
 Db 337 acqehnvplcvpvdnggcytaevkdfvgeyiksadgiarr--lknenklfyggtvrhry 394

Qy 199 QVCEQEDVPVDFHFFYANPDHADLNLVALPRMKTWIRIGLPKVHLSSPKSEQAIRS 258
 | : | | :
 Db 395 pfcwrttdsliykavns-----fvavekvksmlkanesihwtpe-hikqgrfgkwleg 448

Qy 259 HADYVDANFLLERQNGTWNIDFIEAKQDKALLRL--MDELSSIRGVK 306
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
 Db 449 ardwa-----isrnywgtptpiw---rsddgellvigsqlealsgqk 490

RESULT 11

AAB03759
 ID AAB03759 standard; Protein; 234 AA.

AC AAB03759;

DT 04-OCT-2000 (first entry)

DE Human H37 amino acid sequence #2.

XX H37; human; Cdc7 regulatory subunit; cytostatic; proliferative; cancer;
 KW anti-proliferative; replication regulator; stem cell.

OS Homo sapiens.

XX WO200026250-A1.

XX 11-MAY-2000.

XX 01-NOV-1999; 99WO-JP06076.

XX 30-OCT-1998; 98JP-0311408.

XX (NTSC-) JAPAN SCI & TECHNOLOGY CORP.

PA (ARAI/) ARAI K.

PA (MASAI/) MASAI H.

PI Arai K, Masai H;

XX WPI; 2000-365580/31.

XX N-PSDB; AAA53484.

XX Human H37 proteins with a Cdc7 activity regulatory subunit, for
 PT controlling cell replication and cell proliferation, useful in treating
 PT cancers and diseases due to abnormal production of stem cells

XX Claim 2; Page 46-47; 55pp; Japanese.

XX The present sequence represents a human H37 protein sequence. H37 is a
 CC protein with a Cdc7 activity regulatory subunit. The invention relates to
 CC two H37 protein and nucleotide sequences. H37 proteins exhibit
 CC cytostatic, proliferative, anti-proliferative, and cell replication
 CC regulatory activities. The proteins, encoded genes and DNA fragments are
 CC useful in treating cancers and other diseases resulting from abnormal
 CC production of stem cells. Antibodies directed against one of the H37
 CC proteins can be used to inhibit cell proliferation.

XX Sequence 234 AA;

Query Match 5.6%; Score 92.5; DB 21; Length 234;
 Best Local Similarity 31.7%; Pred. No. 0.33;
 Matches 33; Conservative 10; Mismatches 42; Indels 19; Gaps 4;

Qy 227 VALPRMKTWIRIGLPKVHLSSPKS-----EQAIRSHADYVDANFLLEFR 273
 | : | | : : : : : : | | | | : : : : : : : : : : : : : : : :
 Db 105 vspesaytaetshpshdsgsfksptvclsrkglivekaikh-dfipsnslsal 163

Qy 274 QWGTNI-----DFMIEAKOKKALLRMLDELSSIR-GVKRIGGGA 312
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
 Db 164 swgvkllhiddiryieqkkkelyllkksstsvrdggkrvgsga 207

RESULT 12

AAB42482

ID AAB42482 standard; Protein; 674 AA.

AC AAB42482;

XX 08-FEB-2001 (first entry)

XX Human ORFX ORF2246 polypeptide sequence SEQ ID NO:4492.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.

XX Homo sapiens.

XX WO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach M;

XX WPI; 2000-602362/57.

XX N-PSDB; AAC76691.

PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX Claim 11; Page 3676-3677; 5507pp; English.
 PS
 CC AACT74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antiproliferative; antiparkinsonian; nootropic; immunosuppressive;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressive;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antididiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX Sequence 674 AA;

Query Match 5.6%; Score 92.5; DB 21; Length 674;
 Best Local Similarity 31.7%; Pred. No. 1.7;
 Matches 33; Conservative 10; Mismatches 42; Indels 19; Gaps 4;
 QY 227 VALPRMIKTWIRIGLPKVHLSPKS-----EQAIRSHADYVDANFLRLER 273
 Db 105 vspesaytaettspshdsgsfkspdtvclsrqkllvekaikdh-dfipsnlsnal 163
 QY 274 QWGTNI-----DFMTEAKOKKALLRLMDELSSIR-GVKRIGGGA 312
 Db 164 swgkvlhiddirryieqkkellyllkksstsvrdgkgkrgvsga 207

RESULT 13
 AAB03758
 ID AAB03758 standard; Protein; 674 AA.
 XX
 AC AAB03758;
 XX
 DT 04-OCT-2000 (first entry)
 XX
 DE Human H37 amino acid sequence #1.
 XX
 KW H37; human; Cdc7 regulatory subunit; cytostatic; proliferative; cancer;
 KW anti-proliferative; replication regulator; stem cell.
 XX
 OS Homo sapiens.
 XX
 PN W0200026250-A1.
 XX
 PD 11-MAY-2000.
 XX
 PF 01-NOV-1999; 99WO-JP06076.
 XX
 PR 30-OCT-1998; 98JP-0311408.
 XX
 PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
 PA (ARAI/) ARAI K.
 PA (MASA/) MASAI H.
 XX
 PI Arai K, Masai H;
 XX

DR WPL; 2000-365580/31.
 DR N-PSDB; AAA53483.
 XX
 PT Human H37 proteins with a Cdc7 activity regulatory subunit, for
 PT controlling cell replication and cell proliferation, useful in treating
 PT cancers and diseases due to abnormal production of stem cells -
 PS Claim 1; Fig 5; 55pp; Japanese.
 XX
 CC The present sequence represents a human H37 protein sequence. H37 is a
 CC protein with a Cdc7 activity regulatory subunit. The invention relates to
 CC two H37 protein and nucleotide sequences. H37 proteins exhibit
 CC cytostatic, proliferative, anti-proliferative, and cell replication
 CC regulatory activities. The proteins, encoded genes and DNA fragments are
 CC useful in treating cancers and other diseases resulting from abnormal
 CC production of stem cells. Antibodies directed against one of the H37
 CC proteins can be used to inhibit cell proliferation.
 XX Sequence 674 AA;

Query Match 5.6%; Score 92.5; DB 21; Length 674;
 Best Local Similarity 31.7%; Pred. No. 1.7;
 Matches 33; Conservative 10; Mismatches 42; Indels 19; Gaps 4;
 QY 227 VALPRMIKTWIRIGLPKVHLSPKS-----EQAIRSHADYVDANFLRLER 273
 Db 105 vspesaytaettspshdsgsfkspdtvclsrqkllvekaikdh-dfipsnlsnal 163
 QY 274 QWGTNI-----DFMTEAKOKKALLRLMDELSSIR-GVKRIGGGA 312
 Db 164 swgkvlhiddirryieqkkellyllkksstsvrdgkgkrgvsga 207

RESULT 14
 AAW77094
 ID AAW77094 standard; Protein; 522 AA.
 XX
 AC AAW77094;
 XX
 DT 09-NOV-1998 (first entry)
 XX
 DE Sorting nexin 1.
 XX
 KW Human; sorting nexin; intracellular domain; cell surface receptor;
 KW translocation; lysosome; down-regulation; cancer.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 115
 FT /note= "Xaa not specified, encoded by TGT"
 XX
 PN US5804412-A.
 XX
 PD 08-SEP-1998.
 XX
 PF 01-APR-1996; 96US-0625322.
 XX
 PR 01-APR-1996; 96US-0625322.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Cadena DL, Gill GN, Kurten RC;
 XX
 DR WPL; 1998-505652/43.
 DR N-PSDB; AAV48248.
 XX
 PT Nucleic acids encoding sorting nexin proteins - used for recombinant
 PT production and regulating expression of the protein, especially in
 PT treating unregulated cell growth e.g. cancer
 XX
 PS Claim 3; Column 23-28; 25pp; English.

XX Sorting nexins (SNX) have the ability to bind an intracellular domain of
 CC an activated cell surface receptor expressed on a cell and direct
 CC translocation of the receptor to a lysosome in the cell. The nucleic
 CC acids encoding SNX are useful for the down-regulation of cell surface
 CC receptors for degradation when they are no longer required to carry out a
 CC signal. The nucleic acids can be used to express the protein in a cell
 CC where expression is desired e.g. where cancer is caused by up-regulation
 CC of receptors. Antisense constructs can be used to inhibit native
 CC expression of the protein.
 XX Sequence 522 AA;

Query Match 5.5%; Score 91.5; DB 19; Length 522;
 Best Local Similarity 19.3%; Pred. No. 1.4;
 Matches 64; Conservative 54; Mismatches 106; Indels 107; Gaps 16;
 QY 28 YSKLSKTER-----KEALLTVTKANL-RNTWRTLHYITGHGIPLYRFSSSIVPLA 76
 Db 194 yeklseksqngfivpppeksligmtkvkgkedsssaeflekrraalerlyqlriv--- 250
 QY 77 THPDVMDVFTPQKFEIGELVKTHQLRTSPHPNQFTLFTSPKESVTKNAV-----TDM 132
 Db 251 nhptlmldq---pdvreflekeelpavgvtqlslgagllmfknkatdavsaktikmnesdi 307
 QY 133 AYHYRMLEANGIADR-----SVINIHTGGAYGNKDTA--TAOPHONIKOLPOEIKERMTL 185
 Db 308 wfeekleqeveceeqrlrkhlhavvetlrvn---hrkelalntaqfaksiamlg----- 355
 QY 186 ENDDKVTTEETLQVCEQEDVPFVDFHFHYANPD-----DH-----ADLNVALPRM 232
 Db 356 sseentalsalsqlaeveek---ieqlhqeanndfillaellsdyirllaivraafdr 413
 QY 233 IKTWERTGLQPKVHLSSPKSEQAIRSHADYVDANFILLERFRQ-----WGTNIDFMIEAK 286
 Db 414 mktwqr-----wqdaqatlgkreaearllwankpdklqqak 450
 QY 287 QKOKALLRLMDELSSIRGVKRIGGGALOWKS 317
 Db 451 -----del-----lewes 458

RESULT 15
 AAY77979
 ID AAY77979 standard; protein; 558 AA.
 AC AAY77979;
 XX
 DF 14-JUN-2000 (first entry)
 XX
 DE A. thaliana environmental stress tolerance related protein.
 XX
 KW Environmental stress; plant; transgenic plant; anaerobic; flooding; cold;
 KW dehydration; drought; heat stress; salinity; osmotolerance.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO200008187-A2.
 XX
 PD 17-FEB-2000.
 XX
 PF 04-AUG-1999; 99WO-EP05652.
 XX
 PR 04-AUG-1998; 98EP-0202634.
 XX
 PA (VLAAs-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.
 XX
 PI Lee JH, Verbruggen N;
 XX
 DR WPI; 2000-205726/18.
 DR N-PSDB; AAZ98359.
 XX

PT Isolation of polynucleic acids useful for producing transgenic plant by
 PT isolating genes involved in tolerance to environmental stress -
 XX Claim 12; Page 278-282; 312pp; English.
 XX
 CC The invention relates to isolation of coding sequences and/or genes
 CC involved in tolerance to environmental stress in plants. The sequences
 CC (AAZ98305-298365) are useful for producing a transgenic plant having
 CC enhanced tolerance or resistance to environmental stress conditions such
 CC as anaerobic, flooding, cold, dehydration, drought, heat stress or
 CC salinity. This is useful for producing improved yield, growth,
 CC development and productivity under environmental stress conditions, and
 CC also provides growth of crops in areas where they cannot grow without
 CC the induced osmotolerance. Sequences AAY77925-984 represent polypeptide
 CC sequences from A. thaliana that are encoded by the genes involved in
 CC environmental stress tolerance.
 XX Sequence 558 AA;

Query Match 5.5%; Score 91; DB 21; Length 558;
 Best Local Similarity 21.8%; Pred. No. 1.8;
 Matches 53; Conservative 45; Mismatches 91; Indels 54; Gaps 11;
 QY 69 SSSIPLATHPDV-----MMDFVT--PFOKEFREIGELVKTHQLRTSPHPNQFTLFTS 119
 Db 114 srsrvssksnptvdafqgkewekltadpgtrvyleqddfvktmk-eiqqrpnlnlmyk 172
 QY 120 PRESVTKNAVTDMAHYHYRMLEAMGIADRSVINIHGGAYGNKDTATATQFHQNIKOLPOEI 179
 Db 173 dk-----rvmkalgv-----llnvkfggssg-edtemkeaderkepepeme 212
 QY 180 KERMTLENDKDTYTTTEETLQVCEQEDVPF-VPDF-----HHFYANPDHDLNVALPRMI 233
 Db 213 pmeleeeerqkkekalkkegnvaykkdkdgravehytkamelddedisylntr-- 270
 QY 234 KTWERTGLQPKVHLSSPKSEQAIRSHADYVDANFILLERFROMGTNIDFMIEA-KOKDKAL 292
 Db 271 -----aavylemgkyecie-----dcdkavergreirsdfkmiatlrkgsal 315
 QY 293 LRL 295
 Db 316 vkm 318

Search completed: January 15, 2002, 13:50:09
 Job time: 298 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 14:07:07 ; Search time 45.07 seconds
(without alignments)
533.662 Million cell updates/sec

Title: US-09-724-296-36

Perfect score: 3482

Sequence: 1 MPSRKSAALDTPQSEST.....REVKKGVPEVEGEFGD 656

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	348.5	10.0	320	1 YWJD_BACSU	P45864 bacillus su
2	146	4.2	796	1 YFC3_YEAST	P43573 saccharomyc
3	141	4.0	747	1 DIL2_HUMAN	Q9ulw0 homo sapien
4	137	3.9	3256	1 KI67_HUMAN	P46013 homo sapien
5	136.5	3.9	848	1 NFM_MOUSE	P08553 mus musculu
6	135	3.9	793	1 CALD_HUMAN	Q05682 homo sapien
7	134	3.8	2663	1 CENE_HUMAN	Q02224 homo sapien
8	132.5	3.8	915	1 ZDS1_YEAST	P50111 saccharomyc
9	131.5	3.8	4473	1 PLE1_CRIGR	Q9ji55 cricetulus
10	131	3.8	845	1 NFM_RAT	P12839 rattus norv
11	128.5	3.7	2004	1 MO2_HUMAN	Q92794 homo sapien
12	128	3.7	1020	1 NFM_HUMAN	P12036 homo sapien
13	128	3.7	1805	1 NEST_RAT	P21263 rattus norv
14	127.5	3.7	2453	1 NCRI_MOUSE	Q60974 mus musculu
15	127	3.6	3130	1 DPOZ_HUMAN	O60673 homo sapien
16	126.5	3.6	551	1 CALX_PEA	O82709 pisum sativ
17	126.5	3.6	1490	1 CRK7_HUMAN	O9ayv4 homo sapien
18	126	3.6	2375	1 ATRX_HUMAN	P46100 homo sapien
19	124.5	3.6	1960	1 MYH9_HUMAN	P35579 homo sapien
20	124.5	3.6	1982	1 CHDM_DROME	O97159 drosophila
21	123	3.5	2517	1 NCR2_HUMAN	Q9v618 h nuclear r
22	122.5	3.5	679	1 YIS3_YEAST	P40563 saccharomyc
23	122.5	3.5	2774	1 MAPA_RAT	P34926 rattus norv
24	122	3.5	530	1 CALX_ARATH	P29402 arabidopsi
25	122	3.5	915	1 NFM_HUMAN	P07197 homo sapien
26	121.5	3.5	540	1 CALX_HELTU	Q39994 helianthus
27	121.5	3.5	2158	1 MY9B_HUMAN	Q13456 homo sapien
28	121	3.5	1087	1 NFM_MOUSE	P19246 mus musculu
29	121	3.5	1781	1 AKAC_HUMAN	Q02952 homo sapien
30	120.5	3.5	531	1 CALD_RAT	Q62736 rattus norv
31	120	3.4	546	1 CALX_SOYBN	Q39817 glycine max
32	120	3.4	743	1 TFE3_HUMAN	P19532 homo sapien
33	119	3.4	892	1 IF2_CHLTR	O84098 chlamydia t

ALIGNMENTS

RESULT 1

YWJD_BACSU STANDARD; PRT; 320 AA.

AC P45864:

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE HYPOTHETICAL 36.9 KDA PROTEIN IN ACDA 5'REGION.

GN YWJD.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI_TaxID=1423;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RA Glaser P., de la Fuente V., Danchin A.;

RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; 249782; CAA89865.1; -

DR EMBL; 299123; CAB5748.1; -

DR Subtilist; BG11309; YWJD.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 320 AA; 36896 MW; 2FD35A710648E002 CRC64;

Query Match 10.0%; Score 348.5; DB 1; Length 320;

Best Local Similarity 32.0%; Pred. No. 1.9e-14;

Matches 87; Conservative 52; Mismatches 104; Indels 29; Gaps 8;

QY 277 ANARDIVMLCWNEXYGRFLRSLSEMPFPASHP-VHGYKLAPFASEVLAEAGRAAEELG 335

DB 46 ANLRMTLRTHTIIGHGIPLYRFSSTIVPLATHPDVMMDFVTPFOKE-FREIGELVKTHQ 104

QY 336 HRLTTHPQQTQLGSPKRVESAIROLEYHDELILKLPEQQNRDAVMTIHMGQFGD 395

DB 105 LRTSPHPNQFTLTSPKESVTKNAVTDMAHYRMLEANGIADR---SVINHIIGGAYGN 160

QY 396 KAATLERKFNVARLSQSKNRLVLENDVGWTVHDLPLVCEELNIPWLVYHHIHCIFD 455

DB 161 KUTATAQHNIKQLPQETIKERMTLENDKTYTTEETLQVCEQEDVPPVDFHHFYANPD 220

QY 456 PAHLREGTLDISDPKLOBRIANTMKRGKIKOMHYSEP-CDCAVTPRRRKRHR----- 507

DB 221 D-----HADLNVALP-----RMKTWERIGLQPKVHLSSPKSEQAI-----RSHADYVDANF 267

QY 508 --PRVMTLPPCPDMDLMEARKEQAVFELM 537

FT	REPEAT	1366	1477	4.
FT	REPEAT	1487	1598	5.
FT	REPEAT	1608	1720	6.
FT	REPEAT	1730	1842	7.
FT	REPEAT	1851	1964	8.
FT	REPEAT	1974	2086	9.
FT	REPEAT	2096	2204	10.
FT	REPEAT	2214	2326	11.
FT	REPEAT	2335	2447	12.
FT	REPEAT	2457	2569	13.
FT	REPEAT	2579	2689	14.
FT	REPEAT	2699	2808	15.
FT	REPEAT	2818	2928	16.
FT	NP_BIND	3034	3041	ATP (POTENTIAL).
FT	VARSPLIC	136	495	MISSING (IN SHORT ISOFORM).
SQ	SEQUENCE	3256 AA;	358741 MW;	578F8C51BED42517 CRC64;

Query Match	3.98;	Score 137;	DB 1;	Length 3256;
Best Local Similarity	19.98;	Pred. No. 2.4;		
Matches 146;	Conservative 92;	Mismatches 296;	Indels 200;	Gaps 30;

Qy	5	KSAAALD-----TP-QSESSTSTLDSSAPSA-----RNLRRSGRNILQPS	48
Db	2428	KEKALEDLVGFKELQTPGCHTEESMTDDKITEVSKSPQSFKTSRSKQRLKPLV	2487
Qy	49	EKDRDHEKRSGEELAGRMGMKANGCHCLREG-----KEQEGVKMAIEGLARME-RLQR	102
Db	2488	KVMKEEPLAYSKLIT-RTSGTGTQHTPTGTDGSKSIKAFKESPKQIILDPAAVTSRRQL	2546
Qy	103	ATKROKKOLEEDGIPVSVVSRFTAPVHHKSTNAEREAKEPV-----	146
Db	2547	RTKKEKARALEDLVDKELFS-----AGHTEESTIDKNTKIPCKSPPELTDTATSKR	2602
Qy	147	-----LKTHSKDVERAEI-GVDDVWMEPAA-----TNII-----EPEDAQAAE	186
Db	2603	CPKTRPRKEVEELSVERLTQTSQTHHKEPASGDEGIKVLKQRAKKKPNPVEEPS	2662
Qy	187	RGAAAPPVAVNSYPLPWKGRGLVACLTLYLRNAKPPIFSSRTCRMASIVDHRPLQFED	246
Db	2663	RRRPAPKEKAQ-----PLEDLAGTELESETSGHTQESLTAGKATK-----IPCESPLEVD	2715
Qy	247	---EPEHLKPKDKSKPEQDGLGHKFQVELGLANARDIVKMLCWNKYGIRFLRLSSE	303
Db	2716	TTASTKRHLRTVRQVKVEEPSAVKFTQTSQ--ETTDADREPAGEDK-GIKALKESAKQ	2772
Qy	304	FPFASHPVHGKLPAPFASEVLAE-----AGRVAAELGH-----RLTHTP-----GQFT	346
Db	2773	TPAPAAVTSGSRRRPRAPRESAQATIEDLAGFKDPAAGHTTESMTDDKTKIPCKSSPELE	2832
Qy	347	QLGSPRKEVESAIIRDLEVHDELLSLKL-----PQQN	380
Db	2833	DPATSKRRPRTRAQKVEKVEKELLAVGKLQTSGETTHDKPEVGEKGTKAFKQPAKRN	2892
Qy	381	RDVAMIIHMGGO---FGDKAATLERFKRNYARLSQSKCNRLVLNDVGVTVHLLPVCE	437
Db	2893	VDAEDVIGSRQRPAPKEKAQPLEDL-ASFQELSOT-----	2927
Qy	438	ELNIPWLDYHHHNCIFDPAHLR-----GTLDISDP-KLOERIANTWKRR	482
Db	2928	-----PGHTEELANGAASFTSAPKQTPDSGKPLKISRRVLRAPK--	2967
Qy	483	GIKQKMHYSEPCDGAVTPRDRKRPVMT-LPPCPPDMDLMIEAKDKEQAVFELMRTEK	541
Db	2968	-----VEPVGDVVSTRDPVKSQSKNTSLPLP-----FKRGGGKDGSVTGTKRLRC	3014
Qy	542	LCGFEKINDWYDRDDENRAPPKAKKKKGGKRRKTTTDEEAAREPEVDTAADDVKDA	601
Db	3015	MPAPEEIVEELPASKQORVAPRARGKSPEPVVIMKRSRLTSKAKRIEPAE-ELNSNDMKTN	3073
Qy	602	PEGPK---EVPPEE	612
Db	3074	KEHKLODSVPENK	3087

RESULT 5
NFM_MOUSE
ID NFM_MOUSE STANDARD: PRT: 848 AA.
AC P08553; Q61961;
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NEUROFILAMENT TRIPLET M PROTEIN (160 KDA NEUROFILAMENT PROTEIN)
DE (NEUROFILAMENT MEDIUM POLYPEPTIDE) (NF-M).
GN NEF3 OR NEFM OR NFM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87246694; PubMed=3036526;
RA Levy E., Liem R.K.H., D'Eustachio P., Cowan N.J.;
RT "Structure and evolutionary origin of the gene encoding mouse NF-M,
RT the middle-molecular-mass neurofilament protein.";
RL Eur. J. Biochem. 166:71-77(1987).
RN [2]
RP SEQUENCE OF 322-540 FROM N.A.
RX MEDLINE=87158637; PubMed=3103856;
RA Julien J.-P., Meyer D., Flavell D., Hurst J., Grosveld F.;
RT "Cloning and developmental expression of the murine neurofilament
RT gene family.";
RL brain Res. 387:243-250(1986).
CC -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NEM IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; X05640; AAA29127.1; -;
DR EMBL; M20481; AAA39815.1; -;
DR PIR; S00030; S00030.
DR MGD; MGI:97314; Nfm.
DR InterPro; IPR001664; IF.
DR Pfam; PF00038; filament; 1.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Neurone; Phosphorylation;
KW Glycoprotein.
FT INIT_MET 0
FT DOMAIN 1 102 HEAD.
FT DOMAIN 103 410 ROD.
FT DOMAIN 411 848 TAIL.
FT DOMAIN 103 134 COIL 1A.
FT DOMAIN 135 147 LINKER 1.
FT DOMAIN 148 246 COIL 1B.
FT DOMAIN 247 263 LINKER 12.
FT DOMAIN 264 285 COIL 2A.
FT DOMAIN 286 289 LINKER 2.
FT DOMAIN 290 410 COIL 2B.
FT CARBOHYD 47 47 O-LINKED (GLCNAC) (BY SIMILARITY).
FT CARBOHYD 430 430 O-LINKED (GLCNAC) (BY SIMILARITY).

FT	CONFLICT	432	432	S -> F (IN REF. 2).
FT	CONFLICT	539	540	QA -> RR (IN REF. 2).
SEQ	SEQUENCE	848 AA:	95910 MW:	5F251F274D0F13B6 CRC64;
Query Match				
Best Local Similarity 20.1%; Score 136.5; DB 1; Length 848;				
Matches 157; Conservative 104; Mismatches 270; Indels 249; Gaps				
QY	3 SRKSKAAALDTPQSSSSFTSLDSSASPARNLRRSGRNILQPSSEKDRDHEKRSGEEL 62			
Db	53 TRSAVAPRLAYSSAMLSAESLDFSQSSLLNGSGDGYKLSRNEK			
QY	63 AG---RMGCKDANGHCLREGKEQEGVKMAIEGLAR			
Db	104 QCLNDRFAGYTEKVHYL--- <td></td> <td></td> <td></td>			
QY	105 ---KROKKQ--- <td></td> <td></td> <td></td>			
Db	161 MVNHEKAQVQLDSHLEED---IHLRKERF--- <td></td> <td></td> <td></td>			
QY	156 REAETGVDDVVKMEPATNII-----EPEDAADAERGAARPPAVNSSYLPL----- 202			
Db	210 ESSMVKVELDKKVSQLODEVAFLRRNHEEEVADLLAQIQAHSHTIVERKDYLTKDISTALK 269			
QY	203 -----PKMGRUGYACL-----NTYLNRNAKPPI-----FSRQTCR 231			
Db	270 ETRSOLECHSDONMHQAEEW-FKCRYAKLTAEAQNKEAIRSAKEETAERYRQLOQSKTE 328			
QY	232 MASIVDHRHPLO-----FEDEPEHLKANPKDKPEQDEL-GHKFVQELGLANARDI--V 283			
Db	329 LESVNGTKESLERQLSDIEERHNHDLSSYQDTIQQLENLGRGKWEMARHREYQDLLNV 388			
QY	284 KMLCNWCKYGTRELRLSSEMPFASH-----PVHGYKL-----AP----- 318			
Db	389 KMALDIEIAYRKL-LGEETRFSTFGSGTGLPYTHROPSVTISSIKQKTVEAPKLKV 447			
QY	319 ---FASVLAERAGVAAELGHLRLTHFGPQTQLGSPRKVEVESAIRDLEYHDELLSLKL 375			
Db	448 QHKEVEETIET-KVEDEKSEMEEETLTAIAELAAASAKEKEEAKEEPEEAESPVKS 506			
QY	376 PEONRDAVMIHHGGGFGDKAATLRFERKRYARLSQSCKNRLVLENDDVGTVTHDLLPV 435			
Db	507 PEAKEEE-----EGEKEEEEEQESEE-----EEDEG----- 534			
QY	436 CEELNIPWLDYHHNCTCFPAHLRGLTLDISDPKLOERITANTWRKGIKQKMHYSEPCD 495			
Db	535 -----VKSDQA--EEGG-----SEKGSSEKDEGQEEEEEGETAE 568			
QY	496 GAVTPDRRRKHPRVMTLPCPPDMDLMTAEAKDQEAQVAFELMRTKLPGLFEKINDMVPYD 555			
Db	569 GEGEAEAKEEK-----KIEKGVEEYAVKKEIKVEK-P--ERAKSPMPKS 610			
QY	556 RDDENRAPVPVKAPK-----KKGGGKRRTTDEAAAEPEVDTAADVDKDAPEGPK-EVP 609			
Db	611 PVBEVKPKPEAKGAKGEQKEEKEVEKEVTKESPKEEKKEKPKADVADKKKASP 670			
QY	610 EERAMGGPNRYWPLGCPDWL-----KPKKREVKKGKVPVEVEDEG 652			
Db	671 VKEKAV-----BEVITISKSVKVSLEKDTKEKQPQEE---KVKEAKEEG 713			
RESULT 6				
ID	CALD_HUMAN	STANDARD;	PRT;	793 AA.
AC	Q05682; Q13979; Q13978; Q14741; Q14742;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DE	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	CALDESMON (CDM)			
GN	CALD1 OR CDM OR CAD.			
OS	Homo sapiens (human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			

DR	MM: 114213; -	
DR	InterPro: IPR000075; Caldesmon.	
DR	Pfam: PF02029; Caldesmon; 1.	
DR	PRINTS: PR01076; CALDESMON.	
KW	Muscle protein; Actin-binding; Calmodulin-binding; Phosphorylation;	
KW	Repeat; Alternative splicing.	
FT	DOMAIN 319 375	3 X 14 AA TANDEM REPEATS OF E-E-K-R-A- A-E-R-Q-R-I-K.
FT	REPEAT 319 332	1.
FT	REPEAT 333 346	2.
FT	REPEAT 347 360	3.
FT	DOMAIN 26 207	MYOSIN AND CALMODULIN-BINDING (BY SIMILARITY).
FT	DOMAIN 564 621	TROPOMYOSIN-BINDING (POTENTIAL).
FT	DOMAIN 664 674	TROPOMYOSIN-BINDING (POTENTIAL).
FT	DOMAIN 653 686	STRONG ACTIN-BINDING (BY SIMILARITY).
FT	DOMAIN 716 722	CALMODULIN-BINDING (BY SIMILARITY).
FT	DOMAIN 768 793	WEAK ACTIN-BINDING (BY SIMILARITY).
FT	DOMAIN 39 46	POLY-ARG.
FT	DOMAIN 81 86	POLY-THR.
FT	DOMAIN 189 196	POLY-GLU.
FT	DOMAIN 376 379	POLY-GLU.
FT	DOMAIN 540 543	POLY-ARG.
FT	DOMAIN 580 583	POLY-GLU.
FT	DOMAIN 597 600	POLY-GLU.
FT	MOD_RES 724 724	PHOSPHORYLATION (BY CDC2) (BY SIMILARITY).
FT	MOD_RES 730 730	PHOSPHORYLATION (BY CDC2) (BY SIMILARITY).
FT	MOD_RES 753 753	PHOSPHORYLATION (BY CDC2) (BY SIMILARITY).
FT	MOD_RES 759 759	PHOSPHORYLATION (BY CDC2) (BY SIMILARITY).
FT	MOD_RES 789 789	PHOSPHORYLATION (BY CDC2) (BY SIMILARITY).
FT	VARSPLIC 1 24	MODERRRRLRQKREMRLEAER -> MLGSGSHGRRSL AALSO (IN ISOFORM HELA L-CAD I AND ISOFORM HELA L-CAD II).
FT	VARSPLIC 208 436	MISSING (IN ISOFORM HELA L-CAD I AND ISOFORM WI-38 L-CAD I).
FT	VARSPLIC 208 462	MISSING (IN ISOFORM HELA L-CAD II AND ISOFORM WI-38 L-CAD II/1-CAD).
FT	CONFLICT 530 530	V -> M (IN REF. 1).
SQ	SEQUENCE 793 AA; 93250 MW; 2A0DC63D16DD6B5F CRC64;	
Query Match 3.9%; Score 135; DB 1; Length 793;		
Best Local Similarity 20.5%; Pred. No. 0.53;		
Matches 143; Conservative 105; Mismatches 271; Indels 180; Gaps 28;		
QY	34 RNLRRSGRNILQPSSEKDRHDKRSGEELAGRMCKDANGHCLREGKROE-----	83
DB	8 RELURKREEMRLAERITAYQRNDDDEFAARERRRRARQERLRQKQBEESLGQVTDQVE	67
QY	84 -----EGVKMA--IEGLARME-----RRLQRAATKROKK---QLEE 113	
DB	68 VNAQNSVPDEAKTTTNTQVGDDEAFLRLARERROKRLQEALEKEFDPTTD 127	
QY	114 DGIPVS-----VSRFPATPYHKTNAER---EAKPEVLKTHSKOVERAEIGVDV 165	
DB	128 ASLSLPSRRMQNDTAENETTEREESQERYEIEETETVTKSYQKNDWRDAENKKED 187	
QY	166 VKMEPAATNIIPEADAADAAERPPAVNSSYLPPLPKWGLRGVACLTLYLRNAKPPIF 225	
DB	188 KEKE-----EEEEPKNG-----SIGENQEVVWVEETTESQETVVMSLKNQOI 233	
QY	226 SSRTCRM-----ASIVDRHPLOFED-----EPHILKNPKDKSKPEQDELG 267	
DB	234 SSEPKQEEEREQSGDSIEHSHEKMEDEKRAEAEARLEAEERIKAEQDKKIADEPA 293	
QY	268 HKFVQELGLANARDIVKMLCWNEKYGIRFLRLSLSEMFPPASHVHCYKLPAPASEVLAE 327	
DB	294 RIAEAEKAAAOERE-----RRAE-----EREMREE 320	

QY	328 GFVAALGHLRTHTPGQFTQLGSPRKEVVESSAIRDLEYHDELLSLKLPEQNDAVMII	387
DB	321 EKRAAEERQRIKE---EEKRAAEERQRIKEEKRAAEERQ-----RIKEEKRAAEERQ	371
QY	388 HMGQFGDKAAATLIERKRYARLSQSKNRLVLENDVGVTHDLLPVCEELNPMVLDY	447
DB	372 RARAEERK-AKVEQKRN-KOLEE--KKRAMQETKIKGEVE-----QKIEGKWNEK	421
QY	448 HHNNICFPAHLR-----EGT-LDISDPKLOERIANWKRKGIKOK--MHYSEPCDGAVT	499
DB	422 KAOEDKLOTAVLKKGQEEKGTQVQAKREKLQED-KPTFKKEIKDEKIKKDEKPEEYKS	480
QY	500 PRDRKRHRPRVMTLPPCPDMDLMTIEAKDKQOAVFELM---RTFKLPGEFEXINDMVPYDR	556
DB	481 FMDRKKGFTEV-----KSONGEDMTKHLKHTENTFSRPGGRASVD-----	520
QY	557 DQNRPAIPVAPKPKKKKGGKRR--TTDEE-----AAEPVEVDTAADDVKDAPEGPKVEPE	610
DB	521 TREAGAPQVEAGKRLLELRRRRGGETESEFEKLOKQOQEALELELKKREERKKVLE	580
QY	611 EERAMGPPYNRVWPLGCPCEWLKPKKREVKGKVPPEVE 649	
DB	581 EEEQRKQ-----EADRLPEEEERKRLKEEIE 609	
RESULT 7		
ID	CENE HUMAN STANDARD; PRT: 2663 AA.	
AC	Q02224;	
DT	01-JUL-1993 (Rel. 26, Created)	
DT	01-JUL-1993 (Rel. 26, Last sequence update)	
DT	30-MAY-2000 (Rel. 39, Last annotation update)	
DE	CENTROMERIC PROTEIN E (CENP-E PROTEIN).	
GN	CENPE.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=93024922; PubMed=1406971;	
RA	Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;	
RT	"CENP-E is a putative kinetochore motor that accumulates just before	
RT	mitosis";	
RL	Nature 359:536-539(1992).	
RN	[2]	
RP	CHARACTERIZATION.	
RX	MEDLINE=95196755; PubMed=7889940;	
RA	Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;	
RT	"Mitotic HeLa cells contain a CENP-E-associated minus end-directed	
RT	microtubule motor";	
RL	EMBO J. 14:918-926(1995).	
RN	[3]	
RP	CHARACTERIZATION.	
RX	MEDLINE=98437347; PubMed=9763420;	
RA	Chan G.K.T., Schaar B.T., Yen T.J.;	
RT	"Characterization of the kinetochore binding domain of CENP-E reveals	
RT	interactions with the kinetochore proteins CENP-F and hBUBR1.";	
RL	J. Cell Biol. 143:49-63(1998).	
CC	-1- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE	
CC	OF THE CELL CYCLE. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE	
CC	AND/OR SPINDLE ELONGATION.	
CC	-1- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.	
CC	-1- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING	
CC	CONGRESSION. RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS	
CC	QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.	
CC	-1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.	
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

DR	EMBL; Z15005; CAA78727.1; -.
DR	PIR; S28261; S28261.
DR	HSSP; P03069; 1Z1I.
DR	MIM; 117143; -.
DR	InterPro; IPR001752; kinesin.
DR	Pfam; PF00225; kinesin; 1.
DR	PRINTS; PR00380; KINESINHEAVY.
DR	SMART; SM00129; Kisc; 1.
DR	PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR	PROSITE; PS00667; KINESIN_MOTOR_DOMAIN2; 1.
DR	Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW	Cell cycle; Centromere.
FT	DOMAIN 1 335
FT	DOMAIN 336 2471
FT	DOMAIN 2472 2563
FT	NP_BIND 86 93
FT	SEQUENCE 2663 AA; 312087 MW; CEFC13880C8RCB8 CRC64

Query Match 3.8%; Score 134; DB 1; Length 2663;
Best Local Similarity 17.9%; Pred. NO. 2.8;
Matches 143; Conservative 126; Mismatches 280; Indels 252; Gaps 32;

QY	16	SESSFTS	TLDS-----	SAPSPARNLRSGRNI-----	43			
			: : :	: : : :				
Db	467	SESDVFN	TLDTLSEIWN	PATKLLNQENISELNSLRADYNLDV	EQLRTEKEEMEL 526			
		: : :	: : :	: : :				
QY	44	---	LOPSS	KDRD-----	HEKRSFELAGRMWGDCANGHC 75			
		: : :	: : :	: : :				
Db	527	KLUKENDL	DEFALER	KYKQDEMOLIH	EISNLKMLVHREVVYODLENELSSK---VEL 583			
		: : :	: : :	: : :				
QY	76	LREGKQ	-----	EGVKM-----	ATBGLARMERRLQRA-----TKRKKQ 110			
		: : :	: : :	: : :				
Db	584	LREKEDQ	IKKLOEY	IDSOKLENIKM	DLSYESIEDPQMKOTLFD	AEVALDAKRESAF 643		
		: : :	: : :	: : :				
QY	111	LREEDG	IPVPSV	SRPTAPYHHK	STNAEREAK	EPVLKTHSKDYVEREA	EIGVDDVVWKEP 170	
		: : :	: : :	: : :				
Db	644	LRSNLEL	KEKKELAT	---YKOMENDI	OLYOSLEAKKM	QVMDLEKELQSAFNI	ITL--- 700	
		: : :	: : :	: : :				
QY	171	AATNIE	---PDA-----	QDAARGAAR	PAVNSSYL---PLPWKGR	L 208		
		: : :	: : :	: : :				
Db	701	--TSLID	GVKPDLLCN	LEGIKTDLQ	KNKEVEENALREEVILL	SELKSLP-----	753	
		: : :	: : :	: : :				
QY	209	GYACLNTY	LRNAKPI	FSRTRCMAS	IVDHRHPLQ	FEDEPHHLKNKPD	KSKFQDELIGH 268	
		: : :	: : :	: : :				
Db	754	---	SEVERL	KKEIQDKSEELH	ITSEKDKL---FSEVYH	786		
		: : :	: : :	: : :				
QY	269	K---	FVQELG---	LANARDIVK-----	MLCWNEKYGIR 295			
		: : :	: : :	: : :				
Db	787	KESRVQGL	LEEIGTK	KDDLATTQ	SNKYKTDQEFQNF	LHMDFEQYKMWLEENRNQ	E 846	
		: : :	: : :	: : :				
QY	296	FLRLSE	PPFASH	PHVGYKLAP	FASEVLA	EAGRVAAELGHLR	LTHPGQFTQLGSPRKEV 355	
		: : :	: : :	: : :				
Db	847	IVNLS	KEAQKDS-----	SLGAKLT	ELSYK-----TOELQ	EKTRE 881		
		: : :	: : :	: : :				
QY	356	VESATRD	LEYHDELL	SLKLPEQ---NRDAV	MIHMGQFGDKAATL	ERFRNYARLSQSC 414		
		: : :	: : :	: : :				
Db	882	VOERLNE	MBQLKE	OLNDRSP	LOTVEREKT	LITEKLOOTLEEV	VTLTOEKDDLKQLOES- 940	
		: : :	: : :	: : :				
QY	415	KNRLV	DDVGV	THDL---LPVCE	EL-NIPVLDYHHNI-----	CFDPAHL 459		
		: : :	: : :	: : :				
Db	941	---LO	TERDQL	KSIDHTV	NNMINDTQ	EOLRNALES	LKHQETINTL	KSISEEVSRLNH 997
		: : :	: : :	: : :				
QY	460	REGTLD	ISDPK	LOERANT	WRKKG	IKORWHYSEPCD	GAVTPDRDKR	HRPRVNTLPPCPD 519
		: : :	: : :	: : :				
Db	998	ENTGET	KD-EFQ	KWGWG	DKKDL	EAKTQTLTAD---VKDNE	ITEEQORKL	IFSLOEKNE 1054
		: : :	: : :	: : :				
QY	520	MDLMT	EA---KKEQ	AVFEL	MRTFK---LPGFEK	I	NDWVPYRDDN	RAPPVAKPK--- 571

CC -!- FUNCTION: MAY REPRESENT A CHROMATIN-ASSOCIATED ACETYLTRANSFERASE.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.
 CC -!- DISEASE: PARTICIPATES IN A T(8;16)(P11;P13) CHROMOSOMAL
 CC TRANSLOCATION THAT PRODUCES A MOZ-CBP CHIMERA OBSERVED IN THE
 CC M4/M5 SUBTYPE OF ACUTE MYELOID LEUKEMIA.
 CC -!- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.
 CC -!- SIMILARITY: BELONGS TO THE MYST (SAS/MOZ) FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; UA7742; AAC50662.1; --
 CC HSSP; Q50631; 4GB0.
 CC MIM; 601408; --
 CC InterPro; IPR001386; Linker_histone.
 CC InterPro; IPR002717; MOZ_SAS.
 CC InterPro; IPR001965; PHD.
 CC Pfam; PF01853; MOZ_SAS; 1.
 CC Pfam; PF00628; PHD; 2.
 CC SMART; SM00526; H15; 1.
 CC SMART; SM00249; PHD; 2.
 CC Proto-oncogene; Chromosomal translocation; Zinc-finger; Repeat;
 KW Nuclear protein.
 FT ZN_FING 206 256 PHD-TYPE 1.
 FT ZN_FING 259 313 PHD-TYPE 2.
 FT DOMAIN 371 379 POLY-SER.
 FT ZN_FING 538 560 C2HC-TYPE.
 FT DOMAIN 788 801 POLY-GLU.
 FT DOMAIN 989 995 POLY-GLU.
 FT DOMAIN 1019 1026 POLY-ARG.
 FT DOMAIN 1069 1078 POLY-GLU.
 FT DOMAIN 1147 1150 POLY-LYS.
 FT DOMAIN 1221 1242 GLU-RICH.
 FT DOMAIN 1267 1302 GLU-RICH.
 FT DOMAIN 1411 1414 POLY-GLU.
 FT DOMAIN 1593 1597 POLY-SER.
 FT DOMAIN 1643 1704 GLN/PRO-RICH.
 FT DOMAIN 1897 1977 MET-RICH.
 FT SITE 1546 1547 BREAKPOINT FOR TRANSLOCATION TO FORM
 FT MOZ-CBP.
 FT SEQUENCE 2004 AA; 225054 MW; 9FEBBAC3792854BA CRC64;
 Query Match 3.7%; Score 128.5; DB 1; Length 2004;
 Best Local Similarity 20.1%; Pred. No. 4.3;
 Matches 136; Conservative 89; Mismatches 260; Indels 193; Gaps 32;
 QY 47 SSEKDRDHEKRSGEELAGR-----MMGKDANGHCLREGKEQEGVKMAIEGLARMERRL 100
 Db 787 SEEEEEEAEAGEEENEPQOEERLEISVGKSVS-----HENKQD-----SYSVSEKKPEVMA 839
 QY 101 QRATRKQKQ-LEEDGIPVPSVRPTAPYHHKSTNAEER-----EAKPEVLKTHSKDVE 155
 Db 840 PVSSTRLSKQVLPDLSPLANSQPSR--RGRWGRKNKTKTOERFGDKSKLLLEETSSAPQE 897
 QY 156 REAEIGVDVVVKWEPAANTNIEPEDAQDAERGAARPAVANSYLPL-----PWKGR 207
 Db 898 QYGECE-----EKSEATQOYTESBOLVASEE---QPSQDGKPDLPKRRRLSEGVEPWRGQ 950
 QY 208 LGYACLTNLYRNAKPPFSSRTCRMASIVDHRHPLQFDEPEHLLKNPKDKSPQDELG 267
 Db 951 L-----KKSP--EALKCRITE-GSERLPTRYSEGDAVLGRFSSESEEEPEPS 996
 QY 268 HRFVDELGLANARDIVKMLCWNEKYGIRFLRSLSEMFPFASHVGHYKLPAPFASEVLAEA 327
 Db 997 PR-----SSSPILTKPTLKRK-KPF-----LHRR 1020
 QY 328 GRVAAELGHRLTHPGQFTQLGSPKVEVESAIRDLEHYDELLSL---LKLPEQONRDV 384

Db 1021 RRVRRKKHINSV-----VTETLSETTEVLDEPFDSDSERPMPRLPEPTFEIDEEEEE-- 1074
 QY 385 MIIHMGOGFGDKAATLERFKRNY-----ARLSQSKNRLVLEDDVGVTVHDLPLV 435
 Db 1075 -----DEN---ELFPREYFRLLSQDVLRCCSSSKRKSDEDEEDSDADDPTPI 1121
 QY 436 CEELNIPVLDYHHNICFD---PAHLREGTLDISDKLQERIANTW-KRGIKOKMHYS 491
 Db 1122 LKPVSLLRKRDVKNSPLEPDTSTPLKKKGW-----PKGKSRKPIHMKRKRPGKPGFKLS 1176
 QY 492 -----EPCDGAVTPRDRRKRPRVM-----TLPCPPDMDLMIEAKOKEQAVF 534
 Db 1177 REIMPVSTQACVIEPI--VSPKAGRK--PKIASEETVEP-KEDMPUPPEERKEEEMQA 1231
 QY 535 ELMRTFKLPGEFKINDMVP--YDRDNRPAAPPVKKAPKKKGKKRKTDE----- 583
 Db 1232 EAAEEAEEGEEDAAASSEVPAAADSSNSPETETKEPEVEEEEEKPRVSEBORSEEEQQ 1291
 QY 584 --EAAEPEEVDTA-----DDVKAPEGPKVEPEERAMGGPYNRVYVPLGCEEWLK 633
 Db 1292 ELEPEPEEEEDAAATAQNDHDADEDDGHLE-----S 1326
 QY 634 PKKREVKKGVPEEVEDE 651
 Db 1327 TKKKELEEOPTREDVKEE 1344
 RESULT 12
 NFH_HUMAN
 ID NFH_HUMAN STANDARD; PRT; 1020 AA.
 AC P12036;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
 DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H).
 GN NFH OR NFH.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88328981; PubMed=3138108;
 RA Lees J.F., Sheldman P.S., Skuntz S.F., Carden M.J., Lazzarini R.A.;
 RT "The structure and organization of the human heavy neurofilament
 subunit (NF-H) and the gene encoding it."
 RL EMBO J. 7:1947-1955(1988).
 CC -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
 SUBSERVED BY THE TWO SMALLER NF PROTEINS.
 CC -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
 PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
 INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 OF AXONAL CALIBER.
 CC -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
 LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X15306; CAA33366.1; -.

EMBL; X15307; CAA33366.1; JOINED.	DR	EMBL; X15308; CAA33366.1; JOINED.	DR	EMBL; X15309; CAA33366.1; JOINED.	DR	PIR; S00979; QFHUH.	DR	MIN; 162230; -.	DR	InterPro; IPR001664; IF.	DR	Pfam; PF00038; filament; 1.	DR	PROSITE; PS00226; IF; 1.	KW	Intermediate filament; Coiled coil; Neurone; Phosphorylation.	FT	DOMAIN 1 100 HEAD.	FT	DOMAIN 101 413 ROD.	FT	DOMAIN 414 1020 TAIL.	FT	DOMAIN 101 132 COIL 1A.	FT	DOMAIN 133 145 LINKER 1.	FT	DOMAIN 146 244 COIL 1B.	FT	DOMAIN 245 266 LINKER 12.	FT	DOMAIN 267 288 COIL 2A.	FT	DOMAIN 289 292 LINKER 2.	FT	DOMAIN 293 413 COIL 2B.	SQ	SEQUENCE 1020 AA; 111780 MW; 1177C9DCB3DCF1D4 CRC64;																																																			
Query Match 3.7%; Score 128; DB 1; Length 1020;																																																																																									
Best Local Similarity 20.9%; Pred. No. 1.9;																																																																																									
Matches 130; Conservative 84; Mismatches 243; Indels 164; Gaps 30;																																																																																									
QY	45	QPSSEKDRHKEKRSGEELAGRMGMKDANGCHLCRGKEQEGVKMAIEGLARMERLRQAT	104	DB	465	QVTEETVEEKEAKEE-----EGKEEGEGEEAEAG-GEEETKSPPAE	507	QY	105	KRQKQLEEDGIPVPSVVSFRPTAPYHHKSTNAEREKAPV-LKTHSKDVEREAETGVD	163	DB	508	EASPEKE-----AKSPVKEAKSPAEEKS--PEKEEAKSPAEEKSPK-----AKSPAK	555	QY	164	DVVKMEPATNIIEPEDAQAERGA--ARPPAVNSYLLPLPKWGLRGYACLTLYLNA	220	DB	556	EPAKSPPEAKS--PEKEEAKSPAEEKSPAEEKSPAEEKSPAEEKSPA-----EKA	604	QY	221	KPIFSRTRCMASIVDHRHPLQFDEPEHLKKNPKDKSKPEQDGLGHKFVQELGLANAR	280	DB	605	KSPVKE-----EAKSPAEEKSPAEEKSPAEEKS--PEKASPTKEEAKS--PEKAKSPK	657	QY	281	DIVKMLCKNEKIGIRFLRLSSMFPFASHPVHGYKLAP--FASEVLADAGRAVAELGHL	338	DB	658	EEAK-----SPEK-----AKSPVKAEEKSPAEEKSPAEEKSPA-----	688	QY	339	THPGQFTOLGSPKRVESATRDLEYHDELLSLLLKLPQONRDVAMIHMGGQFGDKAA	398	DB	689	-----KSPEK--AKSPVKE-----EAKSPEKAK-----SPVKEEAK	717	QY	399	TLERPK---RNYARLSQCKNRLVLENDVGTWVTHDLLPVGCEELNIPWLDYHHHNCFD	455	DB	718	SPEKAKSPVKEEAKTPEKAKS-----PVKEEAKSP-----PVKEEAKSP-----	748	QY	456	PAHLRE--GTLDISDPKLOERLANTWKRKIGIKQKMHYSEPCDGAVTPRDRKRRPRVMTL	513	DB	749	KAKSPEKATLDVKSPEAKTPAKEARSADKFPPEKAKSPVKEEVKSPKAK-----	800	QY	514	PCPPDMDLMEIAKQEQAVFELMTFKLPGFEKINDMVPYDRDDENRPAPEKAPKPKK	573	DB	801	SPLKADAKAPEKIPKKEEVKSPVKEEKPQEVKVE--PPKAAEEKAAPATP--KTEEKK	857	QY	574	GGRKRTTDEEAAP---BEVDTAADVKDAP--EGPKVPEERAMGPGYNYWPLGCE	629	DB	858	DSKKEEAPKKEAPKPKVEKPEKVEKPKESKVEAKKEAEADKKKVPTEKAPAKVEK	917	QY	630	EWLKPK-KREVKKGVPEEVE	649	DB	918	EDAKPKTEKAVK-KEPDDAK	937	RESULT 13	NEST_RAT

QY 255 KPDKSKEP-----QDELGHKFOE----- 273
 Db 962 LEDESQETFGPLEKENAESLRSAGQDEQKLEQETQOQLTRAVGNEQMAVSPPEKVDPE 1021
 QY 274 -----LGLANARDIVKMLCWNEKYGIRFLR--LSSEMF----- 304
 Db 1022 LKPLGNDQEIARSLGKENGESIVSL-----KEGIETVKSLETEIIEPLETAEDLERKK 1077
 QY 305 -----PFASHPVHGKLPAPFASEVLAEGRAVAAELGHLRTHHPGQFTGLGSPRKVEVES 358
 Db 1078 SIDQELPLWSTEVARETPEPEPPGSLGSV--DENRETLTSLKESESQELSLGKNNVET 1136
 QY 359 AIRDL-----EYHDELLSLK--LPEQNR-----DAV----- 384
 Db 1137 RVDSQOCLQVEGLQEQHESLREVQKQLPSSGQNRWEDVVEGKAVGOEAPLATTGV 1196
 QY 385 -----MIIHGGQFGKATLFRKKNYARLSOSCKNRVLNDDVG--WTVHDLPLVCE 437
 Db 1197 GTEDKAEHLRGOGGEEAAEG-----ELQDIVGEAWSLSSEP--K 1238
 QY 438 ELNIPM-VLDYHHNHCDFPAHLRGCTLDI---SDPKLOER-----IANTWK 480
 Db 1239 EQRVPREALD-----NLEGALEVPVAQSMPEVTERDEDRQAAGEQDSIEVTLG 1287
 QY 481 RKGIQKMHYSEPCDGAVTPRDRKRHRPRVMTLPCCPPDMLM-----IEAKDKE--- 530
 Db 1288 LEAARTGLEQEVGLEDPR---HFAREEAIPLSLGESVKAKTAQGLEGFGKEPKEA 1343
 QY 531 ----QAVFELMRT----FKLPQFEKNDMVPYDR-----DDENRPAPPVKAP--KKK 572
 Db 1344 GALDSGILELPKTSSEALPCOGHESEMEGWEEREAASLETSDHSGSDAPQPRPETED 1403
 QY 573 KGKRRKT-----TD-----EAAPEPEVDTAADDVVDAPB- 603
 Db 1404 EGQAALTAGPKLLEPCSPILPILDAHELQOAGIOAGWOPAGSSEALERVEVEPEF 1463
 QY 604 GKPEVPEERAMGPNRYNWPVLGCEENLKPKKREYKKGKVPVEVEDEGE 653
 Db 1464 GLGEIPE-----GLQDW-----EEGRESEADDLGE 1489

RESULT 14

NCRI_MOUSE
 AC Q60974; Q60812; STANDARD; PRT; 2453 AA.
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NUCLEAR RECEPTOR CO-REPRESSOR 1 (N-COR1) (N-COR) (RETINOID X RECEPTOR INTERACTING PROTEIN 13) (RIP13).
 DE INTERACTING PROTEIN 13 (RIP13).
 GN NCOR1 OR RXRIP13.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
 RC TISSUE=Pituitary;
 RX MEDLINE=96008539; PubMed=7566114;
 RA Hoerlein A.J., Naeae A.M., Heinzel T., Torchia J., Gloss B.,
 RA Kurokawa R., Ryan A., Kamei Y., Soederstrom M., Glass C.K.,
 RA Rosenfeld M.G.;
 RT "Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor.";
 RL Nature 377:397-404(1995).
 RN [2]
 RP SEQUENCE OF 1792-2453 FROM N.A. (ISOFORM LONG).
 RC TISSUE=Liver;
 RX MEDLINE=95280959; PubMed=7760852;
 RA Seol W., Choi H.S., Moore D.D.;
 RT "Isolation of proteins that interact specifically with the retinoid X receptor: two novel orphan receptors.";
 RT

RL Mol. Endocrinol. 9:72-85(1995).
 CC -!- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME
 CC NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS
 CC PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.
 CC -!- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B
 CC AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES
 CC WITH THE THYROID (TR) AND THE RETINOID ACID RECEPTORS (RAR) IN THE
 CC ABSENCE OF LIGAND.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
 CC -!- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT
 CC ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2
 CC AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-
 CC INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION
 CC DOMAINS (ID1 AND ID2).
 CC -!- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED
 CC SEQUENCE REFERRED TO AS THE CORNR BOX. THIS MOTIF IS REQUIRED AND
 CC SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RARS. SEQUENCES
 CC FLANKING THE CORNR BOX DETERMINE NUCLEAR HORMONE RECEPTOR
 CC SPECIFICITY.
 CC -!- SIMILARITY: CONTAINS 1 SANT DOMAIN (COMPOSED OF SANT-A-SANT-B).
 CC -!- SIMILARITY: CONTAINS 2 CORNR BOX.
 CC -!- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS
 CC FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

EMBL: U35312; AAC17125.1; -;
 DR EMBL: U22016; AAC52168.1; -;
 DR MCD: MGI:1349717; Ncor1.
 DR InterPro: IPR001005; Myb_DNA_bind.
 DR Pfam: PF00249; myb_DNA-binding; 2.
 DR SMART: SM00395; SANT; 2.
 DR PROSITE: PS50090; MYB_3; 1.
 DR Nuclear protein; Transcription regulation; DNA-binding; Repressor;
 KW Coiled coil; Alternative splicing
 FT DOMAIN 174 216 COILED COIL (POTENTIAL).
 FT DOMAIN 254 312 INTERACTION WITH SIN3A/B (BY SIMILARITY).
 FT DOMAIN 299 328 COILED COIL (POTENTIAL).
 FT DNA_BIND 437 482 SANT-A (POTENTIAL).
 FT DNA_BIND 624 669 SANT-B (POTENTIAL).
 FT DOMAIN 501 550 COILED COIL (POTENTIAL).
 FT DOMAIN 606 616 PRO-RICH.
 FT DOMAIN 2073 2077 CORNR BOX OF ID1.
 FT DOMAIN 2277 2281 CORNR BOX OF ID2.
 FT DOMAIN 58 64 POLY-GLN.
 FT DOMAIN 593 602 POLY-ALA.
 FT DOMAIN 1044 1047 POLY-PRO.
 FT DOMAIN 1713 1718 POLY-ALA.
 FT DOMAIN 1968 1979 POLY-SER.
 FT VARSPPLIC 2333 2371 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 1952 1952 I -> T (IN REF. 2).
 FT CONFLICT 2090 2090 A -> P (IN REF. 2).
 SQ SEQUENCE 2453 AA; 270640 MW; 52208B40382F7E6A CRC64;

Query Match 3.7%; Score 127.5; DB 1; Length 2453;
 Best Local Similarity 18.6%; Pred. No. 6.3;
 Matches 123; Conservative 80; Mismatches 232; Indels 225; Gaps 28;

QY 3 SRKSKAALDTPOSESTFSTLDDSSAPARNLRSGRNILQPSSEKDRD-----HEX 56
 Db 103 SLESKRPRLE--QVSDSHFQRI--SAAVLPVHTLPEG---LRSSANAKDPAPGVKHEA 155
 QY 57 RSGEELAGRMGMKDANGCHLREGKEQEGVKMAIEGLARMERLQRA-----TKRQKKQ 110

Db	156	PS-SPLSQPCGDDONASPKLSKEE-----LIQSDRVDRDREIAKVEQOILKLKKQQQ	208	RT	cerevisiae REV3.";
Qy	111	LEEDGIPVSVSRFPATYAPHHKSTNAEERAEKPEVLKTHSKDVERAEAGVDDVVKMEP	170	RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
Db	209	LEEEAAKPE-----PEKVPSPVPVEQKHRSITQIYIDENRKKAEAAHKIFEGGLGPKVEL	263	RP	SEQUENCE OF 79-3130 FROM N.A.
Qy	171	AATNIIPEEDAQAARGAARPPAVNSYLPDPWKGRGLGACLTNYTLRNKAKPPIESRTC	230	RA	Roth T., Rasio D., Murakumo Y., Negrini M., Croce C.M., Fishel R.;
Db	264	PLYN--QPSDTK-----VYHENIK-----TNQVM	285	RL	Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
Qy	231	RMASTVDHRHPLQDEPEHHLKNPKDSKEPDELGHKFVQELGLANARDIVKMLCWE	290	RP	SEQUENCE OF 79-3130 FROM N.A.
Db	286	RKLLL-----FFKRNHARKOREQKICORYDL-----MEANEK	320	RC	TISSUE-Testis;
Qy	291	KYGIREFRLSLSEMFPFASHPVHGYKLAPFASEVLAEAGRAVAELGHLRTHHPGQFTOLGS	350	RA	MEDLINE-99126315; PubMed-9925914;
Db	321	K-----VDRIE-----NN	328	RX	Morelli C., Mungall A.J., Negrini M., Barbanti-Brodano G., Croce C.M.;
Qy	351	PRKEVESAIRDLEYHDELLSLLLKLPQONRDVNIIMHGQFGDKKATLFRFRKNVRL	410	RT	"Alternative splicing, genomic structure, and fine chromosome
Db	329	PRKAKESKTREY-YEKOFFEIRKOREQER-----FORVGORGAGLSATARSSEHISEI	383	RT	localization of REV3L."
Qy	411	SOSCKNLRLVLENDVGVTVHDLPLVCEELNIPMVLVDYHHNICF-----DPA-----	457	CC	Cytogenet. Cell Genet. 83:18-20(1998).
Db	384	IDGLSEQ-----ENNEKQMRQLSVIP-----PMFDAQRRVKFTNMNGLMEDPMKVYKD	433	CC	!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =
Qy	458	-----HLREGTLD--ISDPKLOERIANTWKRGIT-----KQKHYSKPCDG	496	CC	N PYROPHOSPHATE + DNA(N).
Db	434	ROFMNVWTDHEKEIFKDFIOHPKFNGLFIASYLERSKSPDCVLYYYITKKNENYK-----	488	CC	!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
Qy	497	AVTPDRKRHRPVMTLPCCPPDMDLMTAKDKQAVFELMRTEKLPGEFINDMVYDR	556	CC	!- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED.
Db	489	ALVRRNYKRRGRNQOI--ARPSQEKVEEEDKA-----EXTEKKEBEKK	533	CC	!- DOMAIN: ITS C-TERMINAL PART COULD SERVE AS THE CATALYTIC DOMAIN
Qy	557	DDENRAPPVPAKPKGGKRRRTTDEAAPEEVDVTTAADVKADEPGKPEVPEERAMG	616	CC	DURING NUCLEOTIDE POLYMERIZATION, WHILE ITS N-TERMINAL PART COULD
Db	534	DDEKDD----DKEDSKETTKKDRTEATAEPEEREQV-----TPRGKRTANSQGRGKG	583	CC	PROVIDE SITES FOR PROTEIN-PROTEIN INTERACTIONS WITH OTHER FACTORS
RESULT	15			CC	DURING TRANSLATION DNA SYNTHESIS.
DPOZ_HUMAN				CC	!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
AC	O60673; O43214;	STANDARD;	PRT: 3130 AA.	CC	!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-79 IS THE INITIATOR.
DT	15-DEC-1998 (Rel. 37, Created)			CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
DT	15-DEC-1998 (Rel. 37, Last sequence update)			CC	between the Swiss Institute of Bioinformatics and the EMBL outstation
DT	20-AUG-2001 (Rel. 40, Last annotation update)			CC	the European Bioinformatics Institute. There are no restrictions on its
DE	DNA POLYMERASE ZETA CATALYTIC SUBUNIT (EC 2.7.7.7) (HREV3).			CC	use by non-profit institutions as long as its content is in no way
GN	REV3L OR POLZ OR REV3.			CC	modified and this statement is not removed. Usage by and for commercial
OS	Homo sapiens (Human).			CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			CC	or send an email to license@sib-sib.ch).
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			CC	EMBL; AF058701; AAC24357.1; -
OX	NCBI_TaxID=9606;			CC	EMBL; AF071798; AAC24009.1; -
RN	[1]			CC	EMBL; AF157476; AAD40184.1; -
RP	SEQUENCE FROM N.A.			CC	EMBL; AF035537; AAB88486.1; -
RC	TISSUE=Fetal brain;			CC	EMBL; AF078695; AAC28460.1; -
RX	MEDLINE-98284025; PubMed-9618506;			CC	MTM; 602776; -
RA	Gibbs P.E.M., McGregor W.G., Maher V.M., Nisson P., Lawrence C.W.;			CC	InterPro: IPR002064; DNA_pol_B.
RT	"A human homolog of the Saccharomyces cerevisiae REV3 gene, which			CC	Pfam: PF00136; DNA_pol_B; 2.
RT	encodes the catalytic subunit of DNA polymerase zeta."			CC	PRINTS; PR00106; DNAPOLB.
RL	Proc. Natl. Acad. Sci. U.S.A. 95:6876-6880(1998).			CC	SMART; SM00486; POLBC; 1.
RN	[2]			CC	PROSITE; PS00116; DNA_POLYMERASE_B; 1.
RP	SEQUENCE FROM N.A.			CC	Transferase; DNA-directed DNA polymerase; DNA replication;
RC	TISSUE=Bone marrow, and Leukocyte;			CC	DNA-binding; DNA repair; Nuclear protein; zinc-finger; Polymorphism.
RX	MEDLINE-99202263; PubMed-10102035;			CC	2N_FING 3042 3057 C4-TYPE (POTENTIAL).
RA	Lin W., Wu X., Wang Z.;			CC	FT 2N_FING 3086 3104 C4-TYPE (POTENTIAL).
RT	"A full-length cDNA of hREV3 is predicted to encode DNA polymerase			CC	FT VARIANT 231 231 Q -> H.
RT	zeta for damage-induced mutagenesis in humans."			CC	FT VARIANT 389 389 S -> T.
RL	Mutat. Res. 433:89-98(1999).			CC	FT VARIANT 1540 1540 K -> E.
RN	[3]			CC	FT VARIANT 2697 2697 S -> T.
RP	SEQUENCE FROM N.A.			CC	FT CONFLICT 237 237 E -> O (IN REF. 4 AND 5).
RC	TISSUE=Bone marrow, and Leukocyte;			CC	FT CONFLICT 1156 1156 Y -> C (IN REF. 4 AND 5).
RX	MEDLINE-99202263; PubMed-10102035;			CC	FT SEQUENCE 3130 AA; 352782 MW; 1C0700900F10BB14 CRC64;
RA	Lin W., Wu X., Wang Z.;			CC	Query Match 3.6%; Score 127; DB 1; Length 3130;
RT	"A full-length cDNA of hREV3 is predicted to encode DNA polymerase			CC	Best Local Similarity 19.1%; Pred. No. 9.3;
RT	zeta for damage-induced mutagenesis in humans."			CC	Matches 151; Conservative 110; Mismatches 270; Indels 260; Gaps 40;
RL	Mutat. Res. 433:89-98(1999).			CC	Qy 14 PQSDSSFTSSLTLDSSAPSPARNLRSGRNTLOPSSEKDRDHEKFRSG----RELAGRMWGK 69
RN	[3]			CC	Db 1493 PRSLSEALISQT---KALSOCKN-----RNSTPSA----FEGOSGLAVIKELLOKROOK 1540
RP	SEQUENCE FROM N.A.			CC	Qy 70 DANGHCLREGKEQEGVKMAIEGLARMERLQKQKOLEEDGIPVPSVSVRFTAP 129
RC	TISSUE=Bone marrow, and Leukocyte;			CC	Db 1541 AQNANTQDPLSNKHQPNKNISG--SLEH--NKANKRTR-----SVTS--PRPK 1583
RX	MEDLINE-99202263; PubMed-10102035;			CC	
RA	Lin W., Wu X., Wang Z.;			CC	
RT	"A full-length cDNA of hREV3 is predicted to encode DNA polymerase			CC	
RT	zeta for damage-induced mutagenesis in humans."			CC	
RL	Mutat. Res. 433:89-98(1999).			CC	
RN	[3]			CC	
RP	SEQUENCE FROM N.A.			CC	
RC	TISSUE=Bone marrow, and Leukocyte;			CC	
RX	MEDLINE-99202263; PubMed-10102035;			CC	
RA	Lin W., Wu X., Wang Z.;			CC	
RT	"A full-length cDNA of hREV3 is predicted to encode DNA polymerase			CC	
RT	zeta for damage-induced mutagenesis in humans."			CC	
RL	Mutat. Res. 433:89-98(1999).			CC	
RN	[3]			CC	
RP	SEQUENCE FROM N.A.			CC	
RC	TISSUE=Bone marrow, and Leukocyte;			CC	
RX	MEDLINE-99202263; PubMed-10102035;			CC	
RA	Lin W., Wu X., Wang Z.;			CC	
RT	"A full-length cDNA of hREV3 is predicted to encode DNA polymerase			CC	
RT	zeta for damage-induced mutagenesis in humans."			CC	
RL	Mutat. Res. 433:89-98(1999).			CC	
RN	[3]			CC	
RP	SEQUENCE FROM N.A.			CC	
RC	TISSUE=Bone marrow, and Leukocyte;			CC	
RX	MEDLINE-99202263; PubMed-10102035;			CC	
RA	Lin W., Wu X., Wang Z.;			CC	
RT	"A full-length cDNA of hREV3 is predicted to encode DNA polymerase			CC	
RT	zeta for damage-induced mutagenesis in humans."			CC	
RL	Mutat. Res. 433:89-98(1999).			CC	
RN	[3]			CC	
RP	SEQUENCE FROM N.A.			CC	
RC	TISSUE=Bone marrow, and Leukocyte;			CC	
RX	MEDLINE-99202263; PubMed-10102035;			CC	
RA	Lin W., Wu X., Wang Z.;			CC	
RT	"A full-length cDNA of hREV3 is predicted to encode DNA polymerase			CC	
RT	zeta for damage-induced mutagenesis in humans."			CC	
RL	Mutat. Res. 433:89-98(1999).			CC	
RN	[3]			CC	
RP	SEQUENCE FROM N.A.			CC	
RC	TISSUE=Bone marrow, and Leukocyte;			CC	
RX	MEDLINE-99202263; PubMed-10102035;			CC	
RA	Lin W., Wu X., Wang Z.;			CC	
RT	"A full-length cDNA of hREV3 is predicted to encode DNA polymerase			CC	
RT	zeta for damage-induced mutagenesis in humans."			CC	
RL	Mutat. Res. 433:89-98(1999).			CC	
RN	[3]			CC	
RP	SEQUENCE FROM N.A.			CC	
RC	TISSUE=Bone marrow, and Leukocyte;			CC	
RX	MEDLINE-99202263; PubMed-10102035;			CC	
RA	Lin W., Wu X., Wang Z.;			CC	
RT	"A full-length cDNA of hREV3 is predicted to encode DNA polymerase			CC	
RT	zeta for damage-induced mutagenesis in humans."			CC	
RL	Mutat. Res. 433:89-98(1999).			CC	
RN	[3]			CC	
RP	SEQUENCE FROM N.A.			CC	
RC	TISSUE=Bone marrow, and Leukocyte;			CC	
RX	MEDLINE-99202263; PubMed-10102035;			CC	
RA	Lin W., Wu X., Wang Z.;			CC	
RT	"A full-length cDNA of hREV3 is predicted to encode DNA polymerase			CC	
RT	zeta for damage-induced mutagenesis in humans."			CC	
RL	Mutat. Res. 433:89-98(1999).			CC	
RN	[3]			CC	
RP	SEQUENCE FROM N.A.			CC	
RC	TISSUE=Bone marrow, and Leukocyte;			CC	
RX	MEDLINE-99202263; PubMed-10102035;			CC	
RA	Lin W., Wu X., Wang Z.;			CC	
RT	"A full-length cDNA of hREV3 is predicted to encode DNA polymerase			CC	
RT	zeta for damage-induced mutagenesis in humans."			CC	
RL	Mutat. Res. 433:89-98(1999).			CC	
RN	[3]			CC	
RP	SEQUENCE FROM N.A.			CC	
RC	TISSUE=Bone marrow, and Leukocyte;			CC	
RX	MEDLINE-99202263; PubMed-10102035;			CC	
RA	Lin W., Wu X., Wang Z.;			CC	
RT	"A full-length cDNA of hREV3 is predicted to encode DNA polymerase			CC	
RT	zeta for damage-induced mutagenesis in humans."			CC	
RL	Mutat. Res. 433:89-98(1999).			CC	
RN	[3]			CC	
RP	SEQUENCE FROM N.A.			CC	
RC	TISSUE=Bone marrow, and Leukocyte;			CC	
RX	MEDLINE-99202263; PubMed-10102035;			CC	
RA	Lin W., Wu X., Wang Z.;			CC	
RT	"A full-length cDNA of hREV3 is predicted to encode DNA polymerase			CC	
RT	zeta for damage-induced mutagenesis in humans."			CC	
RL	Mutat. Res. 433:89-98(1999).			CC	
RN	[3]			CC	
RP	SEQUENCE FROM N.A.			CC	
RC	TISSUE=Bone marrow, and Leukocyte;			CC	
RX	MEDLINE-99202263; PubMed-10102035;			CC	
RA	Lin W., Wu X., Wang Z.;			CC	
RT	"A full-length cDNA of hREV3 is predicted to encode DNA polymerase			CC	
RT	zeta for damage-induced mutagenesis in humans."			CC	
RL	Mutat. Res. 433:89-98(1999).			CC	
RN	[3]			CC	
RP	SEQUENCE FROM N.A.			CC	
RC	TISSUE=Bone marrow, and Leukocyte;			CC	
RX	MEDLINE-99202263; PubMed-10102035;			CC	
RA	Lin W., Wu X., Wang Z.;			CC	
RT	"A full-length cDNA of hREV3 is predicted to encode DNA polymerase			CC	
RT	zeta for damage-induced mutagenesis in humans."			CC	
RL	Mutat. Res. 433:89-98(1999).			CC	
RN	[3]			CC	
RP	SEQUENCE FROM N.A.			CC	
RC	TISSUE=Bone marrow, and Leukocyte;			CC	
RX	MEDLINE-99202263; PubMed-10102035;			CC	
RA	Lin W., Wu X., Wang Z.;			CC	
RT	"A full-length cDNA of hREV3 is predicted to encode DNA polymerase			CC	
RT	zeta for damage-induced mutagenesis in humans."			CC	
RL	Mutat. Res. 433:89-98(1999).			CC	
RN	[3]			CC	
RP	SEQUENCE FROM N.A.			CC	
RC	TISSUE=Bone marrow, and Leukocyte;			CC	
RX	MEDLINE-99202263; PubMed-10102035;			CC	
RA	Lin W., Wu X., Wang Z.;			CC	
RT	"A full-length cDNA of hREV3 is predicted to encode DNA polymerase			CC	
RT	zeta for damage-induced mutagenesis in humans."			CC	
RL	Mutat. Res. 433:89-98(1999).			CC	
RN	[3]			CC	
RP	SEQUENCE FROM N.A.			CC	
RC	TISSUE=Bone marrow, and Leukocyte;			CC	
RX	MEDLINE-99202263; PubMed-10102035;			CC	
RA	Lin W., Wu X., Wang Z.;			CC	
RT	"A full-length cDNA of hREV3 is predicted to encode DNA polymerase			CC	
RT	zeta for damage-induced mutagenesis in humans."			CC	
RL	Mutat. Res. 433:89-98(1999).			CC	
RN	[3]			CC	
RP	SEQUENCE FROM N.A.			CC	
RC	TISSUE=Bone marrow, and Leukocyte;			CC	
RX	MEDLINE-99202263; PubMed-10102035;			CC	
RA	Lin W., Wu X., Wang Z.;			CC	
RT	"A full-length cDNA of hREV3 is predicted to encode DNA polymerase			CC	
RT	zeta for damage-induced mutagenesis in humans."			CC	
RL	Mutat. Res. 433:89-98(1999).			CC	
RN	[3]			CC	
RP	SEQUENCE FROM N.A.			CC	
RC	TISSUE=Bone marrow, and Leukocyte;			CC	
RX	MEDLINE-99202263; PubMed-10102035;			CC	
RA	Lin W., Wu X., Wang Z.;			CC	
RT	"A full-length cDNA of hREV3 is predicted to encode DNA polymerase			CC	
RT	zeta for damage-induced mutagenesis in humans."			CC	
RL	Mutat. Res. 433:89-98(1999).			CC	
RN	[3]			CC	
RP	SEQUENCE FROM N.A.			CC	
RC	TISSUE=Bone marrow, and Leukocyte;			CC	
RX	MEDLINE-99202263; PubMed-10102035;			CC	
RA	Lin W., Wu X., Wang Z.;			CC	
RT	"A full-length cDNA of hREV3 is predicted to encode DNA polymerase			CC	
RT	zeta for damage-induced mutagenesis in humans."			CC	
RL	Mutat. Res. 433:89-98(1999).			CC	
RN	[3]			CC	
RP	SEQUENCE FROM N.A.			CC	
RC	TISSUE=Bone marrow, and Leukocyte;			CC	
RX	MEDLINE-99202263; PubMed-10102035;			CC	
RA	Lin W., Wu X., Wang Z.;			CC	
RT	"A full-length cDNA of hREV3 is predicted to encode DNA polymerase			CC	
RT	zeta for damage-induced mutagenesis in humans."			CC	
RL	Mutat. Res. 433:89-98(1999).			CC	
RN	[3]			CC	
RP	SEQUENCE FROM N.A.			CC	
RC	TISSUE=Bone marrow, and Leukocyte;			CC	
RX	MEDLINE-99202263; PubMed-10102035;			CC	
RA	Lin W., Wu X., Wang Z.;			CC	
RT	"A full-length cDNA of hREV3 is predicted to encode DNA polymerase			CC	
RT	zeta for damage-induced mutagenesis in humans."			CC	
RL	Mutat. Res. 433:89-98(1999).			CC	
RN	[3]			CC	
RP	SEQUENCE FROM N.A.			CC	
RC	TISSUE=Bone marrow, and Leukocyte;			CC	
RX	MEDLINE-99202263; PubMed-10102035;			CC	
RA	Lin W., Wu X., Wang Z.;			CC	
RT	"A full-length cDNA of hREV3 is predicted to encode DNA polymerase			CC	
RT	zeta for damage-induced mutagenesis in humans."			CC	
RL	Mutat. Res. 433:89-98(1999).			CC	
RN	[3]			CC	
RP	SEQUENCE FROM N.A.			CC	
RC	TISSUE=Bone marrow, and Leukocyte;			CC	
RX	MEDLINE-99202263; PubMed-10102035;			CC	
RA	Lin W., Wu X., Wang Z.;			CC	
RT	"A full-length cDNA of hREV3 is predicted to encode				


```

QY 130 YHKSTNAEREAREKPVKTHSKDVERAEI-----GVDDVVVKMEPAATN 174
Db 1584 RTPRETKQKEKIPK--LLKVDLSNLQNSQSDSDSPFSDPGFESCYSLEDS--- 1638
QY 175 IIEPE-----DAQDAERGAARPAVNSSYLP-----LPWKGRGLGYACLNTYLNRNKPPIF 225
Db 1639 -LSPEHNYNFDINTIGOTGFCFSYS-GSOFVPADQNLPOK-----FLSDAVQDILF 1686
QY 226 SSRTCRMASIVDHRHPLQFEDEPEHIL-----KNKP 256
Db 1687 PGOAIEKNEFLSHDN--OKDEDKHHTTDSASWIRSGTLSPEIFEKSTIDSNEENRRHNM 1744
QY 257 DKSKEPDELGHKFVQELGLANARDIVKMLCWNEKYGIRFLRLSSEMFPFASHPVHGYKL 316
Db 1745 KNSFHPLTTRNSIMDSFCVQQAED-----CLSEKSLNRSSVSKEVFLSLPQPNNSDWI 1799
QY 317 APFASEVLAERAGRAVAELGHLRTHPGQFTO-LGSPRKEVWESAIRDLEYHDELLSLLKL 375
Db 1800 -----QGHTRKEMQOSLDSANTSETAILSSPDGELVDVACBDLE-----L 1839
QY 376 PEQQRDAVMIIHMGQFGDKAATLEREKRNRYARLSQCKNRLVLENDVGVTVHDLPLV 435
Db 1840 YVSRNDML-----TPTDSSPRSTSSPSQS-KNGSFTPR-----TANILKPL 1881
QY 436 C-----EELNIPVLDY-----HHNICDPAPHLREGTLDISDPKL--QERIAN----- 477
Db 1882 MSPPSREEI-MATLLDHLSETIYQEPFCSPNPSDVPEKPREIGRLLMVETRLANDLAEF 1940
QY 478 -----TWKR-----KGIKQMHYSEPCDGAVTPRDRKRHRPR 509
Db 1941 EGDFLSLEGLRLMKTAFSAMTONPRPGSPLRSGQGVVYKGNSSNP-----KMVEDKK 1991
QY 510 VMTLP-PCPPDMDLM---IEAKDKEAVFELMRTFKLPGFEKINDMVPYDRDENRPP 565
Db 1992 IVIMPCKCAPSROLVOVWLOAKEEYERSKKLPKT-KPTGVVKSANFNSSVNPDDKPVVP 2050
QY 566 VK-----APKKKGGKRRKRTTDEEAAEFEEVDTAADDVKD-- 600
Db 2051 PKMDVSPCILPTAHTKEDVDNSQIALQAPTTCQSQTASESOMLPP--VASASDPEKDED 2108
QY 601 -----APEGP-----KEVPEERAMGGPYNRVYWP-----LGCEEWLKPCKRE 638
Db 2109 DDNYYISYSSPDSVPVPPWQQPISPDSKALNGD-DRPSSPVEELPSLAFENFLKPIKD 2167
QY 639 VKKGKYPVEEVE 649
Db 2168 IQKSPCSEPOE 2178

```

Search completed: January 15, 2002, 14:07:17
Job time: 946 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:51:04 ; Search time 57.72 Seconds
(without alignments)
255.755 Million cell updates/sec

Title: US-09-724-296-36

Perfect score: 3482

Sequence: 1 MPSRKRAALDTPQSEST.....REVKKGVPEVEGEFGD 656

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	131.5	3.8	1098	4	US-08-923-992A-8
2	128	3.7	1805	1	US-07-853-913-2
3	121	3.5	1780	1	US-08-769-309A-5
4	121	3.5	1780	3	US-08-994-570-5
5	119.5	3.4	1128	4	US-08-923-992A-6
6	118.5	3.4	1164	4	US-08-923-992A-2
7	117.5	3.4	1104	4	US-08-923-992A-4
8	117.5	3.4	1588	5	PCT-US93-07261-11
9	117.5	3.4	1663	5	PCT-US93-07261-16
10	117.5	3.4	1872	1	US-08-188-582-14
11	117.5	3.4	1872	1	US-08-646-715-14
12	117.5	3.4	1893	1	US-08-188-582-11
13	117.5	3.4	1893	1	US-08-646-715-11
14	115	3.3	727	2	US-08-475-844-9
15	115	3.3	727	5	PCT-US95-08429-9
16	114.5	3.3	1618	1	US-07-853-913-4
17	113.5	3.3	1018	1	US-08-072-610-2
18	113.5	3.3	1018	2	US-08-719-822B-2
19	113.5	3.3	1018	4	US-09-092-458-2
20	113.5	3.3	1093	4	US-09-315-793-52
21	113	3.2	542	4	US-08-675-816-6
22	112	3.2	558	1	US-08-285-440-6
23	112	3.2	558	1	US-08-630-349-6
24	112	3.2	1068	4	US-09-085-199B-11
25	110.5	3.2	251	2	US-08-766-738-3
26	109.5	3.1	1213	4	US-09-413-814-79
27	109	3.1	410	2	US-09-018-628-14

28 109 3.1 410 3 US-09-273-378-14 Sequence 14, Appl
29 109 3.1 456 2 US-09-018-628-16 Sequence 16, Appl
30 109 3.1 456 3 US-09-273-378-16 Sequence 16, Appl
31 109 3.1 500 2 US-09-018-628-18 Sequence 18, Appl
32 109 3.1 500 3 US-09-273-378-18 Sequence 18, Appl
33 109 3.1 500 4 US-09-018-635-27 Sequence 27, Appl
34 108 3.1 1838 4 US-09-120-663-2 Sequence 1, Appl
35 107.5 3.1 251 2 US-08-766-738-1 Sequence 4, Appl
36 107 3.1 885 2 US-08-533-306A-4 Sequence 1, Appl
37 107 3.1 885 2 US-08-742-923A-4 Sequence 2, Appl
38 107 3.1 1788 2 US-08-962-284-2 Sequence 2, Appl
39 107 3.1 2237 1 US-08-455-543A-48 Sequence 48, Appl
40 107 3.1 2237 2 US-08-223-305C-48 Sequence 48, Appl
41 107 3.1 2337 3 US-08-713-118-2 Sequence 2, Appl
42 107 3.1 2337 4 US-09-452-007-2 Sequence 2, Appl
43 107 3.1 2339 1 US-08-455-543A-47 Sequence 47, Appl
44 107 3.1 2339 2 US-08-223-305C-47 Sequence 47, Appl
45 106 3.0 532 1 US-08-285-440-5 Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-08-923-992A-8
; Sequence 8, Application US/08923992A
; Patent No. 6280738
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Blake, Milan S.
; TITLE OF INVENTION: No. 6280738-IgA FC Binding Forms of the Group B
; TITLE OF INVENTION: Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,992A
; FILING DATE: 05-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,707
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1098 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-923-992A-8

Query Match 3.8%; Score 131.5; DB 4; Length 1098;
Best Local Similarity 18.4%; Pred. No. 0.0042;
Matches 121; Conservative 94; Mismatches 259; Indels 183; Gaps 27;

QY 13 TPQSESTFTSLDSDAPSARNLRNRSGRNIIQPSSEKDRDHEKRSGEELA--GRMMGKD 70

```
Db 31 TDGNNSSSELETTREIMPTDIDKAVEPVKTAGETSAHTGKREKQLOQWKNLKN 90
Qy 71 ANGHCLREGKQOEGVKMAI-----EGLARMERLQORATK-----RQK 108
Db 91 VDNTIL--SHEQKNEFKTIDETNDSALLELENOFNETNKLHLHKOHEEVEKQKAKQ 148
Qy 109 KOLEDGIPVSVVSRFTAPYHHKSNAE--EREAKFVLKTHSKDVERAEIGVDDVK 167
Db 149 KTLKSDTKVD--LSNIDKELNHQSQEAGITNEDKDSMLK-KIEDIRKQAO-----197
Qy 168 MEPAATNIEPDQAADAAERPAVNSSYLLPWKGRGLGYACLTNYLRNAKPPFPSS 227
Db 198 -----QPKKEDAEVK-----VREELGKLFFS 219
Qy 228 RPCRMAISVDHRHPLQFDEPEHLKPKDKSKEPO--DE-----LGHKFVOELGLA 277
Db 220 TKAGL-----DOEQEHVKKETSSEENTQKVDHYANSLQNAOKSLEELDKA 267
Qy 278 NARDIVKMLCWNKEYGIRFLRLSSMFPFASHPVHGYKLPFASEVLAEGRAVAELGHR 337
Db 268 TTNEQATQV--KNQFLENAOKLKEIQPLIKET--NVKLYKAMSESLEQ---VEKELKH- 318
Qy 338 LTHFGQFTQLGSPKPEVSESAIRLDYHDELLSLKLPEQONRDVAVMLIHMGGQFGDKA 397
Db 319 --NSEANLEDLVAKSEIVREYEGKLNOSKNLPQLQLEEAHSAHKLKOVVE---DFRKKF 373
Qy 398 ATLE-----RFRKRYA-----RLSQCKNRLVLENDVGVTVHDDLPLVCEELNIP 442
Db 374 KTSQVTPKKRLKRLDLAANNENQKIELTVSPENITVYEGEDVKTV-----420
Qy 443 MVLDTYHHNICFDPHAHLRGTLDSID-----PKLOERANTWKPKGKQKMHYSEPCDG 496
Db 421 -----TAKSDSKTTLDFSLDTKYNPSVSDRISTNYKNTNTDNHKT-----A 461
Qy 497 AVTPRRDRKRRPRVMTLPPCPDMDLMIKADKQAVFELMRTFKLPGEKINDMV---P 553
Db 462 EITIKNLNESQTVL-----KAKDDSGNVVE--KTFTITVOKKEKOVKPTP 508
Qy 554 YRDDENRAPPVYKAPKKKGKRRKTTDEEAPEEVDVTAADDVKDAPEGPKEYPE 610
Db 509 EQKDSKTEEKVP--QEPKSNKQLOELIKSAQOQLEKLEKATKELMEOPPEIPSN-PE 563
```

RESULT 2

```
US-07-853-913-2
; Sequence 2, Application US/07853913
; Patent No. 5338839
; GENERAL INFORMATION:
; APPLICANT: McKay, Ronald D.G.
; APPLICANT: Lendahl, Urban
; TITLE OF INVENTION: Nestin Expression As An Indicator of
; TITLE OF INVENTION: Neuroepithelial Tumors
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/853,913
; FILING DATE: 19920319
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/660,412
; FILING DATE: 22-FEB-1991
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/603,803
; FILING DATE: 25-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/201,762
; FILING DATE: 02-JUN-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/180,548
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-4641AAAA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1805 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-853-913-2
```

Query Match 3.7%; Score 128; DB 1; Length 1805;

Best Local Similarity 18.8%; Pred. No. 0.018;

Matches 167; Conservative 102; Mismatches 287; Indels 334; Gaps 41;

```
Qy 13 TPOSESTFSTLDSAPSPARNLRSGRNILOPSSEKORDHEKRSGE--LAGRMGK 69
Db 685 SPEEDOEACRPLQKNOEPLGYEEAEG--ILERLEKESQESLSRSPPEEDQAGRSLOK 743
Qy 70 -----DANGHCLKE--GKEQEGVKMAIEGLARMERRQORATKROKKOLEEDG---115
Db 744 ENOEPLGYEAEADOMLERLLEKESQESLSKSPEN--QRIGKPLERENOKSLRYLEENQETF 802
Qy 116 IPVSVSVSRPTAPYHHKSTNAEERAKFVLKTHSKDVERAEIGV-----DDVVK 167
Db 803 VPLESRNQR----PLRSLEVEEERQIRVPLEKVSQDSLSGLAEENVPRLYLEEDDCIN 858
Qy 168 M-----EPAATNIIPEDAQDAARPAVNSSYLLPWKGRGLGYAC 212
Db 859 KSLLKEDTKHSLGSLDRNGDSIIIPQESQTV--SLRPPPEEDQRI-----903
Qy 213 LNTYLRNAKPPIFSSRTRCMASIV-----DHRHPLQFEDPEPHHLKN 254
Db 904 VNHLEKESQE--FSRSSEEEQVNMERSLEGHENHESLSSVEKEQDMVESQLEKESQDSGKS 961
Qy 255 KPDKSKEP-----QDELGHKFFVOE-----273
Db 962 LEDESQETFGPLEKENAESLSLAGQOEQKLEQETQOTLRVAGNEQMAVSPPEKYDPE 1021
Qy 274 -----LGLANARDIVKMLCWNKYGIRFLR-LSSEMF-----304
Db 1022 LPKPLGNQETIARSIGKENQESLSVSL---KEGIVTKVKSLEIILEPLETAEDLERRK 1077
Qy 305 -----PFASHPVHGYKLPFASEVLAEGRAVAELHRLTTHFGQFTOLGSPKPEVVES 358
Db 1078 SIDTQEPLEWSTEVARETVPEPPDPPGSLGV-DENRETLTSLEKESQELSSLCKNNVET 1136
Qy 359 AIRDL-----EYHDELLSLIK--LPEQONR---DAV-----384
Db 1137 KVESQOQLQVEEGLOEQHOESLREVYKQELPSSGNQORWEDVVEGKAVGQEAPLATTGV 1196
Qy 385 -----MIHMGQFGDKAATLREKRNRYARLSQSKNRLVLENDVGV--WTVHDLPLVCE 437
Db 1197 GTEDKAEHLRLRQOGGEEAAAG-----ELLQDIVGEANSLGSGSEP--K 1238
Qy 438 ELNIPM-VLDYHHHNICFDPHAHLRGTLDI----SDPKLOER-----IANTWK 480
Db 1239 EQRVPAEALD-----NLEGALEVPVAOSMPFVTERDEDRAGAGQDSIEVTLG 1287
Qy 481 RKGIKQKMHYSEPCDCAVTPRDRKRRPRVMTLPPCPDMDLM-----TEAKUKE-----530
```


TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1780 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-994-570-5

Query Match 3.5%; Score 121; DB 3; Length 1780;
Best Local Similarity 18.3%; Pred. No. 0.079;
Matches 134; Conservative 95; Mismatches 283; Indels 220; Gaps 30;

QY 4 RKSAALDTPQ-----SSSTFSSITLSDSAPSPARNLRSGRNILQPSSEKDRHQR 57
DB 185 KDKTERPDITVOLLTVKKDEGEAGAGDHDQPSLGAAGEAASKESEPKQSTERPEETLKR 244
QY 58 --SGEELAGRMGKDANGHCLREGKEQEBEGVKMAIEGLARMRRLQRAATRKQKQLEEDG 115
DB 245 BQSHAEISPPAESQAVECKEGERKE-----KEPSKSAESPT 284
QY 116 IPVP-----SVVSREPTAPYHHKSTNAEREAKPEVLKTHSKDVEREAE-IGVDDVVKMEP 170
DB 285 SPVSETGSTKFKFTQGWAGWRKTSFRKPKDEVEASEKKKEQEPKYDTEDEGKAEV 344
QY 171 AATNIIPEDAQDAEAGAAPPAVNSY--LPLPWKGRIGYACLNLYLNKAPPTFSSR 228
DB 345 ASEKLTAEOAHPQPAESAHEPRLSAEKVELPSEQVS-----GSQGPSEKPK 395
QY 229 TCRMASIVDHRHPL-OPDEPEHLKKNPKDSKE-----PDDELGHKFKVQELG 275
DB 396 APLATEVEDEKIEVHOEVVAEVHVSVEETEQKTEVEETAGSVPAEELVGMDAEPQE 455
QY 276 LANARDIVKMLCWNEKYGIRFLRSLSEMFPFASHPVHGYKLAPFASEVLAEAGRAVAELG 335
DB 456 AEPKELVKL-----KETCVSGEDPTQADLSPDEKVLSPKPPGVVSEV- 499
QY 336 HRLTHPCQFTQLQSPKREKVVESAIROLEYHDELLSLKLPEQONRAVMIHMGQFGD 395
DB 500 --EMLSSQRMKVOGSPLKLTST-----GLKLSGKKQKG-----KRGG--GD 540
QY 396 KAATLERKRYARLSQCKNRLVLENDVGTVDHLLPVCCEELNIPWLDYHHHNICFD 455
DB 541 E-----ESGE-----HTQVPA-----DSP-----D 555
QY 456 PAHLREGTLDISPKLQRIANTWKRKGIKOKMHSPECDGA----- 497
DB 556 SOEQKQGESSASSPEPEI--TCLEGLAEVQODGAEAGATSDGKKRREGVTPWASF 613
QY 498 --VTPDRKRHRPRVMTLPPCPDMDLMEAKDEQAVFELMRFTKLPGEK-INDMV 554
DB 614 KMTTPK-KRVRRP-----SESDREDEL-DKVKSAATLSSTESTASMOEE 655
QY 555 DRDDENRPPVYKAPK-----KGGKRRRTTDEE-----AAEPEV 591
DB 656 MKGSVEEPK--EEPKRKVDTSVSWAELICVSSKKRARRSSDDEEGPKAMGGDHQA 713
QY 592 DTAADVDKADPG-----PREVPEERAMGPNRVVWPLGCEEWLKPCK-----RE 638
DB 714 DEAGDKETGTGILAGSQEHDPGQSSSPQAGSPTBGE-----GVSTWESFKRLVTPRK 769
QY 639 VKKGVKPEVED 650
DB 770 KSKSLEEKSED 761

RESULT 5
US-08-923-992A-6
Sequence 6, Application US/08923992A

Patent No. 6280738
GENERAL INFORMATION:
APPLICANT: Rai, Joseph Y.
APPLICANT: Blake, Milan S.
TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B
TITLE OF INVENTION: Streptococcal Beta Antigens
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/923,992A
FILING DATE: 05-SEP-1997
CLASSIFICATION: 536
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 60/024,707
FILING DATE: 06-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1128 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-923-992A-6

Query Match 3.4%; Score 119.5; DB 4; Length 1128;
Best Local Similarity 18.8%; Pred. No. 0.054;
Matches 123; Conservative 95; Mismatches 266; Indels 169; Gaps 29;

QY 13 TPQSESTFSTLSDSAPSPARNLRSGRNILQPSSEKDRHQRSGEELA---GRM--- 66
DB 26 TDGNNSSSSSELETTTMEIPTTDIKKA---VEPV-----EKTAGETSATDTCKREKQ 74
QY 67 -----MGKDANGHCLREGKEQEBGVKMAI-----EGLARMERLQRA TK---RQKK 109
DB 75 LOOKNNLKNVDNTIL--SHEQKNEFTKIDETNDSADALLELENOFNETNRLLIHQHE 132
QY 110 OLEEDGIPVSVVSREPTAPYHHKSTNAEREAKPEVLKTHSKDVEREAEIGVDDVVKM- 168
DB 133 EVELD-----KKAKQCKTLKQSDTKVDLSNIDKELNHQKSPVEKMA 173
QY 169 EP-AATN-----IIEPEDAQDAEAGAAPPAVNSYLPPLPWKGRGLGVACINTYLRNAK 221
DB 174 EPKGTINEDKDSMLKIEDIRKQAQADKKEDA-----EVKVRREL 214
QY 222 PTFSSERTCMASIVDHRHPLQFQDEPEHLKKNPKDSKEPQDELGHKFTVOELGLANARD 281
DB 215 GKLFSTKAGLDOEI-HEH-VKKTSSSEENTOKVDHYANSLOKLAQKSELEEDKATTTNE 272
QY 282 IVKMLCWNKYGIRFLRSLSEMFPFASHPVHGYKLAPFASEVLAEAGRAVAELGHLRTH 341
DB 273 QATQV---KNQFLENAOKLKEMOPLIKET--NVKLYKAMSESLQ---VEKELKH---NS 321
QY 342 PGQFTQLGSPRKEVESAIRDLEYHDELLSLKLPEQONRAVMIHMGQGGDKAATLE 401
DB 322 EANLEDLVAKSKELVREYEGKLNQSKNLPDELKOLEBEAHSKLQVVVE---DFRKKFKTSE 378

QY 402 -----RFSKNYA-----RLSQCKNRLVLENDVCGTVDHLLPVCEELNIPWVLD 446
 Db 379 QVTKRVRKRDLANENNQQKIETLVSPENITVYEGEDVKFTV----- 421
 QY 447 YHHNICFDPFAHREGTLDIS-----PKLQRIANTWKRKGIGIKOMHYSEPCDGAATP 500
 Db 422 -----TAKSDSKTTLDLFDSDLLTKYNSVSDRISTNYKNTDNIKI-----AEITI 466
 QY 501 RDRKRHRPRVWTLPPCPDMDLMEAKDEQAVFELMRTEKLPGEKINDMV---PYDRD 557
 Db 467 KNLKLNESQTVTL-----KAKDDSGNVVE---KTFTITVQKKEKQVPKTPEQXD 513
 QY 558 DENRPAVPVAKPKKKGGKRRKTTDEEAPPEEVDTAADVDKDAPEGKPEVPE 610
 Db 514 SKTEKVP-QEPKSNKQNLQELIKSAQOELEKLEKRAIKELMEOPEIPSN-PE 564
 RESULT 6
 US-08-923-992A-2
 ; Sequence 2, Application US/08923992A
 ; Patent No. 6280738
 ; GENERAL INFORMATION:
 ; APPLICANT: Tai, Joseph Y.
 ; APPLICANT: Blake, Milan S.
 ; TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B
 ; TITLE OF INVENTION: Streptococcal Beta Antigens
 ; NUMBER OF SEQUENCES: 34
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
 ; STREET: 1100 New York Avenue, N.W., Suite 600
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/923,992A
 ; FILING DATE: 05-SEP-1997
 ; CLASSIFICATION: 536
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/024,707
 ; FILING DATE: 06-SEP-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Esmond, Robert W.
 ; REGISTRATION NUMBER: 32,893
 ; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 371-2600
 ; TELEFAX: (202) 371-2540
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1164 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 US-08-923-992A-2
 Query Match 3.4%; Score 118.5; DB 4; Length 1164;
 Best Local Similarity 18.4%; Pred. No. 0.069;
 Matches 122; Conservative 96; Mismatches 256; Indels 189; Gaps 29;
 QY 13 TQSSSTFSSTLDSASAPARNLRSGNIIQPSSEKDRDHEKRSGEELA---GRM--- 66
 Db 62 TDQGNSSSELETTKMEIPTTDIKKA---VEPV-----EKTAGETSATDTGKREKQ 110
 QY 67 -----MGKDANGHCLREGKEOEGSVKMAI-----EGLARMERRLORATK-----ROKK 109

Db 111 LQWKNNLKNLDVNTJL--SHEQKNFEKTKIDETNDSALLLENOFNETNRLLLHIKQHE 168
 QY 110 QLEEDGIPVPSVVSREPTAPYHUHSTNAEREAKBPVLKTHSKDVEREAEIGVDVVKM- 168
 Db 169 EVEKD-----KKAQOKTLKQSDTKVDLSNLDKELNHOKKSOVERMA 209
 QY 169 -EPAATN-----IIEPDAQDAEAGARPVAVNSYLLPLPWKGRGLGYACILNTYLRNAK 221
 Db 210 EQKGITNEDKSMKKKIEDIRKQAOQADKEDA-----EVKVRREL 250
 QY 222 PPIESSRTRMASIVDHRHPLOFEDPEPEHLKKNKDPKSKPEQ--DE-----IGHKPV 271
 Db 251 GKLFSTSTKAGL-----DQEIOEHVKKKETSSEENTQKVDEHYANSQNLAQKSL 298
 QY 272 QELGLANARDIVKMLCWNEKYGIRFLRLSSEMPFASHPVHGKYLAPFASEVLAEAGRVA 331
 Db 299 BELDKATTNEQATV---KNQFLANAOKLKEIOLIKET--NVKLYKAMSESLEQ---VE 350
 QY 332 AELGHLRLTHPGQFTQLGSPKREVVESAIROLEYHDELLSLKLPEQONRDVAMLIHMG 391
 Db 351 KELKH---NSEANLEDLVAKSKEIVREYEGKLNQSKNLPKQLEEAHSLKQVVE--- 404
 QY 392 QFGDKAATLE-----REKRYA-----RLSQCKNRLVLENDVCGTVDHLLPV 436
 Db 405 DFRKKFKTSEQVTPKKRVKRDIAANENNQQKIETLVSPENITVYEGEDVKFTV----- 457
 QY 437 EELNIPWLDYHHNICFDPFAHREGTLDIS-----PKLQRIANTWKRKGIGIKOMHY 490
 Db 458 -----TAKSDSKTTLDLFDSDLLTKYNSVSDRISTNYKNTDNIKI--- 497
 QY 491 SEPCDGAATPDRRKRHRPRVWTLPPCPDMDLMEAKDEQAVFELMRTEKLPGEKIND 550
 Db 498 -----AEITIKNLKLNESQTVTL-----KAKDDSGNVVE---KTFTITVQKKEK 539
 QY 551 MV---PYDRDDENRPAVPVAKPKKKGGKRRKTTDEEAPPEEVDTAADVDKDAPEGKPE 607
 Db 540 QVPKTEOKDSKTEKVP-QEPKSNKQNLQELIKSAQOELEKLEKRAIKELMEOPEIPSN 598
 QY 608 VPE 610
 Db 599 -PE 600
 RESULT 7
 US-08-923-992A-4
 ; Sequence 4, Application US/08923992A
 ; Patent No. 6280738
 ; GENERAL INFORMATION:
 ; APPLICANT: Tai, Joseph Y.
 ; APPLICANT: Blake, Milan S.
 ; TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B
 ; TITLE OF INVENTION: Streptococcal Beta Antigens
 ; NUMBER OF SEQUENCES: 34
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
 ; STREET: 1100 New York Avenue, N.W., Suite 600
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/923,992A
 ; FILING DATE: 05-SEP-1997
 ; CLASSIFICATION: 536
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/024,707
 ; FILING DATE: 06-SEP-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Esmond, Robert W.
 ; REGISTRATION NUMBER: 32,893
 ; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 371-2600
 ; TELEFAX: (202) 371-2540
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1164 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 US-08-923-992A-2

us-09-724-296-36.rai

```

NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSER: John H. C. Blasdale
STREET: One Giralda Farms
CITY: Madison
STATE: New Jersey
COUNTRY: USA
ZIP: 07940-1000
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 6.0.5
SOFTWARE: Microsoft Word 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07261
FILING DATE: 19930805
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/927,531
FILING DATE: 07-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Blasdale, John H. C.
REGISTRATION NUMBER: 31,895
REFERENCE/DOC/KET NUMBER: DX0288K
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-822-7398
TELEFAX: 201-822-7039
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1588 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-07261-11

```

```

; FILING DATE: 19930805
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/927,531
; FILING DATE: 07-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Blasgale, John H. C.
; REGISTRATION NUMBER: 31,895
; REFERENCE/DOCKET NUMBER: DX0388K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-822-7398
; TELEFAX: 201-822-7039
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1588 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US93-07261-11

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: BlasGale, John H. C.
; REGISTRATION NUMBER: 31,895
; REFERENCE/DOCKET NUMBER: DX0288K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-822-7398
; TELEFAX: 201-822-7039
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1588 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-07261-11

Query Match 3.4%; Score 117.5; DB 5; Length 1588;
Best Local Similarity 19.4%; Pred. No. 0.14;
Matches 132; Conservative 108; Mismatches 247; Indels 193; Gaps

QY 41 RNTLPSSPKDNRDHEKRSGEELAGRMWGDANGCHLRGCKEQEGVKMAIEGLARMERLL 100
Db 18 RNALK---LKKLKEQKNDAQAKADLTKE-----QDSSEKSLKEKVNGEALKEKEN 68
QY 101 QRATKQKQKQLEEDGIPVSVSRPTAPYIIHKSTNAEREAKEPV-----LKTHSKD 153
Db 69 KETLK--KKELE-----NOKEKEKNKTKDNDEALKNGND 103
QY 154 -----VEREAETGVDDVVKMEPAATNIIPEDAQDAABERGAARPAVNSSYLPWPWKR 207
Db 104 KDPKKTVPKPPESVEKDLKEMELCKEFTK-QHLKDYERKEKRRNNIILRSRDLKRL 162
QY 208 LGVACINTYLRNAKPPTFSRTCRMASTVDHRPL-----QPDE-----PEHH 251
Db 163 EQLEKLNQLESAINELKERASRRPMVMKMQGMKDEWDWIKKYDDQEAENGTGDEE 222
QY 252 LKNKPQKSFE-----PQDELG-----HKFVQELGLANARDIVKMLCWNE 290
Db 223 IKDKGDGYEEIVETKFGMRKENGALGEIDVEERYEKRYLYKEDGEGDLKDVEEKL---E 279
QY 291 KYGIRFLRLSSEMPFASHPVHCYKLAPFASEVLAEAGRAAELGHLRTHTPGOFTOLGS 350
Db 280 ETGYGF----REKEP-----TTRILVRKR-----NK 302
QY 351 PRKEVESAIRDL-----EYHDELLSLIKLPEQONRDVAMIHMGGOGFGKAATLERFKRN 406
Db 303 EQKKLEDREKKLJAAEPDDE--KKIKDSDSKVVPVNKKNSFPDKFRAPDKRTM 360
QY 407 YARLSQCKNRVLNDDVCGWTVHDLPLPVCEELINIPWLDYHHNICFPFAHLREGTLDI 466
Db 361 FYRLSELFP--IVPRKDNE-----LAVCGD-SMDSKVNKKLKSFTNPFKRRRNKL-- 408
QY 467 SDPKLQF--KIANTWKRKGIKQKMHYSEPCDGAVTPRDRKRKRPVMTLPPCPDMDLMI 524

```

RESULT 8
PCT-US93-07261-11
; Sequence 11, Application PC/TUS9307261
; GENERAL INFORMATION:
; TITLE OF INVENTION: PEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF

Db 409 KERKMOELHKKKNYK---YOKLLEREKRE---NPDGEPLNTPETIHVIRP-----SDLMD 458
 QY 525 EAKDKEQA-VFELMRTFKLPGFEKINDWVYDRDDENRPPVKKPKKKGG-KRRKRTTD 582
 Db 459 KGENKSAGHPFKYQPTKGLKEYEESHVSKDYQLEHE-----PPTKLPYEKGHVSREYQLD 514
 QY 583 EAAAEPEVDTAADDVKADPEGPKVEPEERAMGGPNRYVWPLGCEEWLKPKKREVKKG 642
 Db 515 HEP-----PTKLPEYK-----GHVSREY-----QLDNEVRDELPEYK 549
 QY 643 KVPPE--VEDEG-----EFD 655
 Db 550 HVSREYQLDNEGPTLKEYD 569

RESULT 9

PCT-US93-07261-16
 ; Sequence 16, Application PC/TUS9307261
 ; GENERAL INFORMATION:
 ; TITLE OF INVENTION: PFEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
 ; NUMBER OF SEQUENCES: 23
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: John H. C. Blasdale
 ; STREET: One Giralda Farms
 ; CITY: Madison
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07940-1000
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: Apple Macintosh
 ; OPERATING SYSTEM: Macintosh 6.0.5
 ; SOFTWARE: Microsoft Word 5.1a
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US93/07261
 ; FILING DATE: 19930805
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/927,531
 ; FILING DATE: 07-AUG-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Blasdale, John H. C.
 ; REGISTRATION NUMBER: 31,895
 ; REFERENCE/DOCKET NUMBER: DX0288K
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 201-822-7398
 ; TELEFAX: 201-822-7039
 ; INFORMATION FOR SEQ ID NO: 16:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1663 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; ORIGINAL SOURCE:
 ; ORGANISM: Plasmodium falciparum
 ; STRAIN: Malayan Camp
 ; PCT-US93-07261-16

Query Match 3.4%; Score 117.5; DB 5; Length 1663;
 Best Local Similarity 19.4%; Pred. NO. 0.15;
 Matches 132; Conservative 108; Mismatches 247; Indels 193; Gaps 33;

QY 41 RNTLOPSSSEKDRDHEKRSSEELAGRMGMKDANGHCLRECKEEOEGVKMAIEGLARMERRL 100
 Db 18 RNALK---EKKLKEQKNDAQAADLTAKES-----QDSSEKSLKEKVGNEALKEKEN 68
 QY 101 QRATKQKQKQLEBDGIPVSVVVSREFTAPYHHKSTNABERAKEPV-----LKTHSKD 153
 Db 69 KETLK--KKELE-----NQEKEKNKIKNDNEALKNKND 103
 QY 154 -----VEREALGVDDVVKMPEPATNIIPEDAQDAAGARPAPVNSSYLPPLPWGR 207
 Db 104 KODKKIVPKKPSVEKDLKEMELKEFEIK-OHLKDYERKEKRRNWLRLSRDLKRLREI 162

QY 208 LGYACILTYLRNAKPIFFSSRTCRMASIVDRHRPL-----OFEDE-----PEHH 251
 Db 163 EOELKINAQLESAINELKERRASRPMMVKMORGKMKDEVLEWIKKYDDQAEKNTKDEE 222
 QY 252 LKNKPDKSKK-----PDDELG-----HKFVQELGLANARDIVKMLCWNE 290
 Db 223 IKDKGDGYEEIVETKEYGMRNALGELDEYBERYKRYLKDGECDLKDVEEKL---E 279
 QY 291 KYGIRFLRLSSEMPFASHPVHGYKLAPFASVLAFAAGRAVAELGHRLLTHPGQTOQLGS 350
 Db 280 ETGYCF---REKFP-----TTRILVKKRR-----NK 302
 QY 351 PRKEVVESAIRDL----EYHDELLSLKLPQOONDAYMI IHMGQFGDGAATLERFKRN 406
 Db 303 EQKLLKEDKEKKLIAAEEPDE--KKIKLSDDDKVVVNVKNKSSFPDKFRAPDKKRTM 360
 QY 407 YARLSQSCNRLVLENDVDVGTVDHLLPVCYELNIPWLDYHHHINICFDPHAHLREGTLDI 466
 Db 361 FYRSELFP--IVPRKDR-----LAVCGD-SMDSKVNGKKLKSTFNPFKRRRNLK-- 408
 QY 467 SDPKLQE--RIANTWKRKGIKOKMHVSEPCGAVTPDRRRKRRPRVMTLPPCPPDMDLMI 524
 Db 409 KERKMOELHKKKNYK---YOKLLEREKRE---NPDGEPLNTPETIHVIRP-----SDLMD 458
 QY 525 EAKDKEQA-VFELMRTFKLPGFEKINDWVYDRDDENRPPVKKPKKKGG-KRRKRTTD 582
 Db 459 KGENKSAGHPFKYQPTKGLKEYEESHVSKDYQLEHE-----PPTKLPYEKGHVSREYQLD 514
 QY 583 EAAAEPEVDTAADDVKADPEGPKVEPEERAMGGPNRYVWPLGCEEWLKPKKREVKKG 642
 Db 515 HEP-----PTKLPEYK-----GHVSREY-----QLDNEVRDELPEYK 549
 QY 643 KVPPE--VEDEG-----EFD 655
 Db 550 HVSREYQLDNEGPTLKEYD 569

RESULT 10

US-08-188-582-14
 ; Sequence 14, Application US/08188582
 ; Patent No. 5534410
 ; GENERAL INFORMATION:
 ; APPLICANT: Tjian, Robert
 ; APPLICANT: Comai, Lucio
 ; APPLICANT: Dynlacht, Brian D.
 ; APPLICANT: Hoey, Timothy
 ; APPLICANT: Ruppert, Siegfried
 ; APPLICANT: Tanese, Naoko
 ; APPLICANT: Wang, Edith
 ; APPLICANT: Weinzierl, Robert O.J.
 ; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
 ; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
 ; NUMBER OF SEQUENCES: 36
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: FLEHR, ROHBACH, TEST, ALBRITTON & HERBERT
 ; STREET: 4 Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94111-4187
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/188,582
 ; FILING DATE: 28-JAN-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Osman, Richard A.
 ; REGISTRATION NUMBER: 36,627

Qy	84	EGVKMAIEGLARMERRIORATKROKKO--LEEDGIPVP-----SVSRPPTAPYHHKS	134
Dz	79	: : : : : : K E R T Y O R R R G G D A C H L P Q N Q T D G G V V Q V - - - - -	135
Dz	4	DAVEAIVESEFTFKGKERITYORRRGGDEACHLPQNQTDGGEVVQV- --NSS	57
Qy	135	TNAEERAKBPVL---KTHSKD--VEREAETGVDD-----VVKME-	169
Dz	58	VQVMYMQLDPTLLOMKTEVNEGTVAPEAAAVDDTOITLVQVNMBEQPINIGELQLVQ	117
Qy	170	-----PAATNIIEPEDA--ODAEERGARPANVSYSYLPLP----	203
Dz	118	: : : : : : : : : : : : :	177
Qy	204	-W-----KRLGY-----ACNLTY-----	216
Dz	178	LEQELPPQEDPSWKDPDYOPPAKKTTKTKKSLRYTEBGKDVDVSIVYDFEEBQQEGLL	237
Qy	217	-----LRNKPDP-----TFSSRTCRMASIVDRHRIPLQFDEPEH--	250
Dz	238	SEVNAEKVVGNMKPKPKTKIKKGKVKTFCCLCSYTCPRRSLLD-RHMKSHTDERPHKC	296
Qy	251	-----HLKNPKDKSKPEDELGHKFVOEELCIANARDIVMCLCWNER	291

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:50:00 ; Search time 120.01 seconds
(without alignments)
404.900 Million cell updates/sec

Title: US-09-724-296-36

Perfect score: 3482

Sequence: 1 MPSRKSKAAALDTPQSEST.....REVKKGVPEVEDEGEFDG 656

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_1101.*
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3472	99.7	656	17 AAW01618	Neurospora crassa
2	3292	94.5	626	21 AAY44502	N. crassa delta228
3	3292	94.5	626	21 AAY44503	B. subtilis delta2
4	1011.5	29.0	828	21 AAY44498	GST signal peptide
5	963	27.7	600	21 AAY44500	GST signal peptide
6	960	27.6	371	21 AAY44499	S. pombe delta228-
7	262.5	7.5	294	21 AAY44505	D. radiodurans del
8	160.5	4.6	1341	21 AAY85657	Human Acinus L pro
9	143.5	4.1	914	22 AAB94420	Human protein sequ
10	139	4.0	747	21 AAB19002	A human microtubul
11	139	4.0	747	21 AAB29444	Human foetal liver

12	137	3.9	3256	21	AAV50976	Human cell cycle p
13	136	3.9	2665	22	AAAI4533	Peptide #967 encod
14	136	3.9	2665	22	AAAI4533	Peptide #987 encod
15	136	3.9	2665	22	AAAI4533	Peptide #941 encod
16	135	3.9	1590	22	AAAB73488	Mouse Rim2, a nove
17	135	3.9	3266	21	AAAB2491	Human ORFX ORF2255
18	134.5	3.9	1117	22	AAAB39632	Human polypeptide
19	134	3.8	2663	22	AAAB39097	Human polypeptide
20	134	3.8	2688	22	AAAB39097	Human polypeptide
21	133.5	3.8	583	21	AAAB5659	Human Acinus S' pr
22	133.5	3.8	980	21	AAAB18294	Plasmodium falcipa
23	131.5	3.8	1129	21	AAAB4462	Amino acid sequenc
24	129.5	3.7	1596	18	AAAB31347	Rat tumour suppress
25	128	3.7	617	22	AAAB16458	Peptide #2892 enco
26	128	3.7	617	22	AAAB16458	Peptide #2869 enco
27	128	3.7	1805	13	AAAB27204	Rat nestin. Rattu
28	128	3.7	1805	15	AAAB60126	Rat nestin. protein
29	128	3.7	2819	22	AAAB35408	Human 07CG27 gene
30	127.5	3.7	2453	21	AAAB12454	HNRCR protein sequ
31	126.5	3.6	1272	22	AAAB4881	Murine protein, SE
32	126.5	3.6	1272	22	AAAB4883	Murine protein, SE
33	126.5	3.6	1490	22	AAAB5644	Novel protein kina
34	126.5	3.6	2243	22	AAAB4884	Murine protein, SE
35	125.5	3.6	1072	22	AAAG70871	C. albicans apoptos
36	125.5	3.6	2441	21	AAAB18161	Plasmodium falcipa
37	125	3.6	1382	18	AAAB31867	Human metastasis-a
38	123.5	3.5	567	20	AAAB88788	Polypeptide fragme
39	122.5	3.5	2518	21	AAAB40574	Human ORFX ORF338
40	122	3.5	530	21	AAAG26284	Arabidopsis thalia
41	122	3.5	530	21	AAAG46611	Arabidopsis thalia
42	122	3.5	567	21	AAAG46610	Arabidopsis thalia
43	121.5	3.5	568	21	AAAB85658	Human Acinus S' pro
44	121.5	3.5	2096	21	AAAB41592	Human ORFX ORF1356
45	121	3.5	1780	19	AAW53863	Human gravin polyp

ALIGNMENTS

RESULT 1

AAW01618
ID AAW01618 standard; Protein: 656 AA

XX AAW01618

XX 20-AUG-1997 (first entry)

XX Neurospora crassa DNA repair enzyme.

XX DNA repair enzyme; cyclobutane-type dimer; (6-4)-bound product;

XX exposure; UV light; ultraviolet; suntan cream; prevention; treatment;

XX skin cancer.

XX Neurospora crassa.

XX JP08266275-A.

XX 15-OCT-1996.

XX 29-MAR-1995; 95JP-0094137.

XX 29-MAR-1995; 95JP-0094137.

XX (SHTS) SHISEIDO CO LTD.

XX WPI; 1996-512663/51.

XX N-PSDB; AAT58286.

XX DNA repair enzyme and related DNA - specifically recognises

XX cyclobutane-type dimer and (6-4)-bound product induced by exposure

XX to UV light

XX Disclosure; Page 10-13; 18pp; Japanese.

XX The present sequence shows a DNA repair enzyme which specifically
CC recognises a cyclobutane-type dimer and (6-4)-bound product induced,
CC respectively from TT and TC sequences in DNA, by exposure to UV light.
CC The enzyme cleaves these sequences at the 5'-side. The DNA repair enzyme
CC can be used in, e.g. suntan cream for prevention and treatment of skin
CC cancer. Conventional endonuclease acts on either the cyclobutane-type
CC dimer or the (6-4)-bound product in damaged DNA, while the enzyme
CC of the invention can act on both of them.

XX Sequence 656 AA;

Query Match	99.7%;	Score 3472;	DB 17;	Length 656;
Best Local Similarity	99.7%;	Pred. No. 1.5e-295;		
Matches 654;	Conservative 1;	Mismatches 1;	Indels 0;	Gaps 0;
QY	1	MPSRKSAAALDTPQSESTFSSTLSSAPSPARNLRSGRNILQPSSEKDRDHEKRSGE	60	
Db	1	mpskskaaaltpqsesstfsstlssapsarnlrrsgnrlqpsssekdrdhekrsg	60	
QY	61	ELAGRMGKDANGCHLREGKEQEGVKMAIEGLARMERLRQATKROKKOLEEDGIPVPS	120	
Db	61	elagrmgkdangchclregkeqegvkmaieglarmerrlqratkrqkkleedgipvps	120	
QY	121	VYSRPTAPYHHKSTNAEREAKEPVLTHTSKDVEREAIEIGVDDVVKMEPAATNIIEPD	180	
Db	121	vysrftapyhhkstnaereakepylthskdkvereaeigvddvkvmeapaatniiepd	180	
QY	181	AQDAERGAARPAVNSSYLPPLPWKGRIGYACLNLYLNAPKPPFSSTTCRMASIVDHRH	240	
Db	181	aqdaaergaarppavnsyylpplwkgriyacylnlylnskppfssrtcrmasivdhrh	240	
QY	241	PLQFDEPEHLLKNPKDKSEPDQLGHKFVOELGLANARDIVKMLCWNKYGIRFLRLS	300	
Db	241	plqfedepehlhknpkdksepdqlghkfvoelglanardivkmlcwnkgyirflrls	300	
QY	301	SEMPFPASHPVHGYKLAFFASEVLAEGRAVAAELGHRLTTHPGQFTQLGSPRKEVVEAI	360	
Db	301	semfpashpvhgyklapfasevlaeagrvaaelghrltthpgqftqlgspkrkevvesai	360	
QY	361	RDLEYHDELLSLKLPEQONRAVMIIMHGQFGDKAATLRFKRYARLSQCKNRLVL	420	
Db	361	rdleyhdeillsklpeqonrdavmiihmgoqfgdkaatlerfkryarlsqcknrlvl	420	
QY	421	ENDDVGVTHDLIPVCEELNIPMLVDYHHNTCFDPAHLREGTLDISPKQERIANTWK	480	
Db	421	enddvgtvhdllpvceelnipmvlidyhnhntcfdpahlregtldisdpkqeriantwk	480	
QY	481	RKGIKOKMHYSEPCDGAFTPRDRKRHRPRVMTLPCCPPDMDLMEAKDKEQAVFELMRTF	540	
Db	481	rkgikokmysepcdgavtprdrkrhrprvmtlppccppdmdlmeakdkegavfelmrft	540	
QY	541	KLPQFEKINDMVPYDRDNRNRPAPPVAPKPKKKGKRKTDTDEAAPEEVDTAADDVKD	600	
Db	541	klpqfekindmvpdydrdnrnrpappvapkpkkgkrkttddeaaapeevdtaaddvkd	600	
QY	601	APGPKVPEERAMGGPNRVYWPVLCGEEWLKPKKREYKKGKVPVEDEGEFDG	656	
Db	601	apegpkvpeeramgppnrvywpvlgceewlkpkkrevkkgkvpvevedegefdg	656	

RESULT 2
ID AAY44502
ID AAY44502 standard; Protein: 626 AA.
XX AAY44502;
AC AAY44502;
XX 27-MAR-2000 (first entry)
DT
XX N. crassa delta228-UV damage endonuclease.
XX Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;

KW glutathione-S-transferase signal peptide; uvel+ gene product;
KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
KW basic site; aptatinum diaduct; mismatched nucleotide pairing;
KW nucleotide alkylation; skin cancer.
XX Neurospora crassa.
XX WO9963828-A1.
XX 16-DEC-1999.
XX 08-JUN-1999; 99WO-US12910.
XX 08-JUN-1998; 98US-0088521.
XX 18-MAY-1999; 99US-0134752.
XX (UYEM-) UNIV EMORY.
XX Boetsch PW, Kaur B, Avery AM;
XX WPI; 2000-116417/10.
XX A new truncated ultraviolet damage endonuclease for treatment of skin
XX cancers -
XX Claim 16; Page 59; 133pp; English.
XX The present sequence is delta228-UV damage endonuclease from N. crassa.
XX Delta228-UVDE is a 288 amino acid deletion of the N-terminal of the
XX uvel+ gene product. This is expressed in frame with a GST leader sequence
XX to generate a fusion protein. This provides stable endonuclease fragments
XX for cleaving a double-stranded DNA molecule that has a distorted
XX structure resulting from UV radiation damage, a photoproduct, an abasic
XX site, mismatched nucleotide pairing, aptatinum diaduct, an intercalated
XX molecule or alkylation of a nucleotide. Uvelp can be used in compositions
XX for internal or topical application and as a therapeutic agent for skin
XX cancers.

QY	1	MPSRKSAAALDTPQSESTFSSTLSSAPSPARNLRSGRNILQPSSEKDRDHEKRSGE	60	
Db	1	mpskskaaaltpqsesstfsstlssapsarnlrrsgnrlqpsssekdrdhekrsg	60	
QY	61	ELAGRMGKDANGCHLREGKEQEGVKMAIEGLARMERLRQATKROKKOLEEDGIPVPS	120	
Db	61	elagrmgkdangchclregkeqegvkmaieglarmerrlqratkrqkkleedgipvps	120	
QY	121	VYSRPTAPYHHKSTNAEREAKEPVLTHTSKDVEREAIEIGVDDVVKMEPAATNIIEPD	180	
Db	121	vysrftapyhhkstnaereakepylthskdkvereaeigvddvkvmeapaatniiepd	180	
QY	181	AQDAERGAARPAVNSSYLPPLPWKGRIGYACLNLYLNAPKPPFSSTTCRMASIVDHRH	240	
Db	181	aqdaaergaarppavnsyylpplwkgriyacylnlylnskppfssrtcrmasivdhrh	240	
QY	241	PLQFDEPEHLLKNPKDKSEPDQLGHKFVOELGLANARDIVKMLCWNKYGIRFLRLS	300	
Db	241	plqfedepehlhknpkdksepdqlghkfvoelglanardivkmlcwnkgyirflrls	288	
QY	301	SEMPFPASHPVHGYKLAFFASEVLAEGRAVAAELGHRLTTHPGQFTQLGSPRKEVVEAI	360	
Db	289	---fpashpvhgyklapfasevlaeagrvaaelghrltthpgqftqlgspkrkevvesai	345	
QY	361	RDLEYHDELLSLKLPEQONRAVMIIMHGQFGDKAATLRFKRYARLSQCKNRLVL	420	
Db	346	rdleyhdeillsklpeqonrdavmiihmgoqfgdkaatlerfkryarlsqcknrlvl	405	

QY 421 ENDDVGWTVHDLPLVPCBELNIPMWLDYHHHNCIFDPAHLREGTLDSDPKLOBRIANTWK 480
 Db 406 enddvgwtvhdllpvceelnipmvdlyhhnncifdpahiregldldspklqeriantwk 465
 QY 481 RKGTKQKMHYSEPCDGAVTDRRKIRPRVMTLPCCPPMDLMEAKDKEQAVFELMRTF 540
 Db 466 rkgtkqkmyhsepcdgavtdrrkrhrprvmtlpccppmdlmeakdkeqavfelmrtf 525
 QY 541 KLPGFEKINDMVPYDRDENRPPVKKPKKKGGKRRKRTTDEAAPEEVDTAADDVKD 600
 Db 526 klpgfekindmvpdydrdenrppvpkkpkkkggkrkrtrtdeaaapeev----- 576
 QY 601 APEGKPEVPEERAMGGPNRVYWPGLGCEEWLKPKKREVKKGKVPPEVEDEGEFDG 656
 Db 577 -----evpeeeramgppnrvywpplgceewlkpkkrevkkgkvpvevedegefdg 626

RESULT 3
 AAY44503
 ID AAY44503 standard; Protein; 626 AA.
 XX AAY44503;
 XX 27-MAR-2000 (first entry)
 DT
 XX B. subtilis delta228-UV damage endonuclease.
 DE Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;
 KW glutathione-S-transferase signal peptide; uvell+ gene product;
 KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
 KW abasic site; platinum diaduct; mismatched nucleotide pairing;
 KW nucleotide alkylation; skin cancer.
 XX
 OS Bacillus subtilis.
 DE
 XX W09963828-A1.
 PD 16-DEC-1999.
 XX
 PF 08-JUN-1999; 99WO-US12910.
 XX
 PR 08-JUN-1998; 98US-0089521.
 PR 18-MAY-1999; 99US-0134752.
 XX
 PA (UYEM-) UNIV EMORY.
 XX
 XX Doetsch PW, Kaur B, Avery AM;
 XX WPI; 2000-116417/10.
 DR
 XX A new truncated ultraviolet damage endonuclease for treatment of skin
 PT cancers -
 PT
 PS Claim 16; Page 59; 133pp; English.
 XX
 CC The present sequence is delta228-UV damage endonuclease from B. subtilis.
 CC Delta228-UVDE is a 288 amino acid deletion of the N-terminal of the
 CC uvell+ gene product. This is expressed in frame with a GST leader sequence
 CC to generate a fusion protein. This provides stable endonuclease fragments
 CC for cleaving a double-stranded DNA molecule that has a distorted
 CC structure resulting from UV radiation damage, a photoproduct, an abasic
 CC site, mismatched nucleotide pairing, platinum diaduct, an intercalated
 CC molecule or alkylation of a nucleotide. Uvell can be used in compositions
 CC for internal or topical application and as a therapeutic agent for skin
 CC cancers.
 XX
 SQ Sequence 626 AA;

Query Match 94.5%; Score 3292; DB 21; Length 626;
 Best Local Similarity 95.4%; Pred. No. 8.5e-280;
 Matches 626; Conservative 0; Mismatches 0; Indels 30; Gaps 2;

QY 1 MPSSKSKAAALDTPOSESTYSSSTLDSAPSARNLRSGRNLOPSSSEKDRDHEKRSGE 60
 Db 1 mpsrkskaaaltdtpsesstssstldssapsarnlrrsgrnilqpssekdrtdehrsge 60
 QY 61 ELAGRMGKDGANGHCLREGKEQEGVKMAIEGLARMERRLORATRKOKKOLEEDGIPVPS 120
 Db 61 elagrmmgkdanghclregkeqeegvknaiieglaemerrlgratrkqkgleedgipvps 120
 QY 121 VYSRPTAPYHHKSTNAEEREAKEPVLKTHSKDVEREAEIGVDDVYVKMEPAATNIIEPD 180
 Db 121 vvsrptapyhhkstnaeereakepvltkthskdvereaeigvddvvykmeapaatniiepd 180
 QY 181 AQDAAEERGAARPAVNSYLLPLPWKGRGLGYACLNITYLNAAKPPIPSSRTCMASIVDRH 240
 Db 181 aqdaaergaarppavnsyllplpwkgrlgyaclnitylnakppifssrtcrmasivdrh 240
 QY 241 PLQFEDEPEHHLKPKDKSKEPQDELGHKFVQELCLANARDIVKMLWNEKYGIRFLRLS 300
 Db 241 plqfedepehlkpkdkskkepqlghkfvqelclanardivkmlw----- 288
 QY 301 SEMFPFASHPVHYGKLAPFASVLAEGRAVAELGHRLLTTHPGQFTQLGSPRKEVESAI 360
 Db 289 ---fpashpvhgylapfasevlaeagrvaaelghrltthpgqftqlgsprkevvesai 345
 QY 361 RDLEYHDELLSLJLLPEQONRDVAMIHMGGQFGDKAATLERFKRNRYARLSQSKNRLVL 420
 Db 346 rdleyhdelisllkpeqonrdavmihmggqfgdkaatlerfkryarlsqsknrlvl 405
 QY 421 ENDDVGWTVHDLPLVPCBELNIPMWLDYHHHNCIFDPAHLREGTLDSDPKLOBRIANTWK 480
 Db 406 enddvgwtvhdllpvceelnipmvdlyhhnncifdpahiregldldspklqeriantwk 465
 QY 481 RKGTKQKMHYSEPCDGAVTDRRKIRPRVMTLPCCPPMDLMEAKDKEQAVFELMRTF 540
 Db 466 rkgtkqkmyhsepcdgavtdrrkrhrprvmtlpccppmdlmeakdkeqavfelmrtf 525
 QY 541 KLPGFEKINDMVPYDRDENRPPVKKPKKKGGKRRKRTTDEAAPEEVDTAADDVKD 600
 Db 526 klpgfekindmvpdydrdenrppvpkkpkkkggkrkrtrtdeaaapeev----- 576
 QY 601 APEGKPEVPEERAMGGPNRVYWPGLGCEEWLKPKKREVKKGKVPPEVEDEGEFDG 656
 Db 577 -----evpeeeramgppnrvywpplgceewlkpkkrevkkgkvpvevedegefdg 626

RESULT 4
 AAY44498
 ID AAY44498 standard; Protein; 828 AA.
 XX AAY44498;
 XX 27-MAR-2000 (first entry)
 DT
 XX GST signal peptide and S. pombe UVDE fusion protein.
 DE GST signal peptide; ultraviolet damage endonuclease; UVDE; Uvelp;
 KW UV irradiation; DNA damage; UV radiation damage; fusion protein;
 KW skin cancer; glutathione-S-transferase.
 XX
 OS Schizosaccharomyces pombe.
 DE Synthetic.
 XX
 XX Key Location/Qualifiers
 FT Peptide 1..229 /label= GST_signal_peptide
 FT Protein 230..828 /note= "S. pombe UVDE"
 FT Misc-difference 11 /note= "Encoded by aaaa"
 FT Misc-difference 85 /note= "Encoded by gtt"
 FT Misc-difference 147 /note= "Encoded by gtt"

KW glutathione-S-transferase signal peptide; uvrl+ gene product;
 KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
 KW abasic site; aplatinum diaduct; mismatched nucleotide pairing;
 KW nucleotide alkylation; skin cancer.
 XX
 OS Dienococcus radiodurans.
 XX
 PN WO9963828-A1.
 XX
 PD 16-DEC-1999.
 XX
 XX 08-JUN-1999; 99WO-US12910.
 XX
 PF 08-JUN-1998; 98US-0088521.
 PR
 PR 18-MAY-1999; 99US-0134752.
 XX
 XX (UYEM-) UNIV EMORY.
 PA
 PA Doetsch PW, Kaur B, Avery AM;
 PI
 PI WPI: 2000-116417/10.
 DR
 XX A new truncated ultraviolet damage endonuclease for treatment of skin
 PT cancers -
 PT
 PS Claim 16; Page 60; 133pp; English.
 XX
 CC The present sequence is delta228-UV damage endonuclease from
 CC D. radiodurans. Delta228-UVDE is a 288 amino acid deletion of the
 CC N-terminal of the uvrl+ gene product. This is expressed in frame with a
 CC GST leader sequence to generate a fusion protein. This provides stable
 CC endonuclease fragments for cleaving a double-stranded DNA molecule that
 CC has a distorted structure resulting from UV radiation damage, a
 CC photoproduct, an abasic site, mismatched nucleotide pairing, aplatinum
 CC diaduct, an intercalated molecule or alkylation of a nucleotide. Uvrlp
 CC can be used in compositions for internal or topical application and as a
 CC therapeutic agent for skin cancers.
 XX
 SQ Sequence 294 AA:
 Query Match 7.5%; Score 262.5; DB 21; Length 294;
 Best Local Similarity 32.2%; Pred. No. 1.1e-14;
 Matches 85; Conservative 38; Mismatches 102; Indels 39; Gaps 13;
 QY 292 YGIRFLRLSSEFP---FASHPVHGKYLAPFASEVLAEGRAVAELGHLRTHPCQFTQL 348
 Db 56 hdirlyrlssifpmlldlagddtgaavithlapqil-eaghaftdagvrlmhpqfivl 114
 QY 349 GSPRKEVYESATRDLEYHDEL---LSILKLPQQNRDAVMIIH-MGQFCGDKAATLERF 403
 Db 115 nsdrpevressvramsharvmdgigltaripwn-----lllhgkggrgaelaali--- 166
 QY 404 KRNYARLSQSCNRNLVLENDVGVTVHDLPLVCEELNIPVMVDYHHNIGCFDPAHLREGT 463
 Db 167 ---pdlpdpvrlriglenderayspaelliceatgtplvfadahhvv-----hdkl 215
 QY 464 LDISPDKLOE---RANTWKRKGIKQKMHYSEPCDCAVTPRDRKRHRPVMTLPCCPDM 520
 Db 216 pqgedpsvrewlratatqppe-qgvvhlsnglieg---pqdirr-hshliadffsayadv 270
 QY 521 -DLMIEAKDKQAV--FELMRTFK 541
 Db 271 pqieveakgkeaaialrlmapfk 294
 RESULT 8
 ID AAY85657
 AC AAY85657 standard; Protein; 1341 AA.
 AC AAY85657;
 XX
 DT 12-FEB-2001 (first entry)

XX Human Acinus L protein sequence.
 DE
 KW Chromatin aggregation induction; human; Acinus; AIDS; herpes; diabetes;
 KW Alzheimer's disease; Parkinson's disease; cardiac infarction;
 KW brain infarction; adenovirus infection; viral hepatitis.
 XX
 OS Homo sapiens.
 XX
 PN W02000061743-A1.
 XX
 PD 19-OCT-2000.
 XX
 PF 07-APR-2000; 2000WO-JP02254.
 XX
 PR 09-APR-1999; 99JP-0103317.
 XX
 PA (SHIO) SHIONOGI & CO LTD.
 XX
 PI Sahara S, Eguchi Y, Tsujimoto Y;
 PI WPI: 2000-665130/64.
 DR N-PSDB; AAC61196.
 DR
 XX A novel polypeptide which induces chromatin aggregation but is not
 PT involved in DNA fragmentation used for the regulation of apoptosis and
 PT screening of potential apoptosis inhibitors -
 PT
 PS Disclosure; Page 60-67; 99pp; Japanese.
 XX
 CC This invention relates to a polypeptide which induces chromatin
 CC aggregation. The polypeptide is a fragment of the human Acinus protein.
 CC Included in the invention is a nucleotide sequence encoding the
 CC polypeptide, antisense oligonucleotides, antibodies recognising the
 CC polypeptide sequence, and apoptosis regulating agents containing the
 CC polypeptide, nucleotide sequence, and sense or antisense
 CC oligonucleotides. Also included in the invention is a method for
 CC screening compounds for their activity as chromatin aggregation
 CC inhibitors, regulators or promoters. The polypeptide induces chromatin
 CC aggregation within the cell nucleus without inducing DNA fragmentation.
 CC The polypeptide can be used in the investigation and treatment of
 CC diseases in which apoptosis is implicated, including AIDS, Alzheimer's
 CC disease, Parkinson's disease, cardiac or brain infarction, herpes or
 CC adenovirus infection, diabetes and viral hepatitis. The present sequence
 CC represents the human Acinus protein.
 XX
 SQ Sequence 1341 AA:
 Query Match 4.6%; Score 160.5; DB 21; Length 1341;
 Best Local Similarity 19.6%; Pred. No. 8.8e-05;
 Matches 153; Conservative 104; Mismatches 315; Indels 209; Gaps 29;
 QY 1 MPRKRKKAALDTPQSESTFSTLDSAPSPARNLRSGRNILQPSSEKDRD-----H 54
 Db 565 lpshrlikgsa---dsstsrsssssssssrssrpsdssgrshslrskqrdaqrth 620
 QY 55 EKRSGEELAG-----RMWGRDANGHCLR-----EGKEQDEGVKM 88
 Db 621 anprgrpkmdgrstseersrstrssasnstkslspgvsrdsstsyetdkdpssgqev 680
 QY 89 ATEGLARM-----ERRLQKATKROKKQLEEDGIPVPSVVSVRPPTAPYHHKSTNAERE- 141
 Db 681 atppvplqlqvcepkertststsssvqarrlsqpesaekhvtqlqpergspkceaeap 740
 QY 142 --AKEP-----VLKTHSKD-----VERAEITGVD-----DVVKMEPAATNIEPED 180
 Db 741 paatqpqlsetqshlpeserihhtveekvumdsenrpendv----pepmpiadqv 796
 QY 181 AQDAAEARGAARPZAVNSSYLPFWKGRLE-----GYACINTLYLRNAKP----- 222
 Db 797 snddrpegsvdeekesslpksfkrrkksvsvsatkgvpagnsdteggqgqrkrwagsta 856


```

QY 103 ATKROKQLEEDGIPVSVSRFPTAPVHHKSTNAEREAKEPV----- 146
Db 2547 rtrkekaraleldvofkeifs-----apgteesntdkntkipcksppltdtatckr 2602
QY 147 -----LKTHSKDVERAEI-GVDDVVRMEPAA-----TNII-----EPEDAQDAAE 186
Db 2603 cpktrprkeveelsaverlqtsgqstthkepasgdegikvlkqakkkpnpveeeps 2662
QY 187 RCAARPPAVNSYLLPLPKWGRGLVACLTLYLRNAKPPITFSSRTCTMASIVDHRHPLQFED 246
Db 2663 rrrprapkeaq---plidlagltelsetsghtgesitagkatk---ipcesppllevd 2715
QY 247 ---EPEHHLKKNPKDSKPEQDELGHKFVQELGLANARDIVKMLCWNKYGIRFLRLSSEM 303
Db 2716 ttastkrhlrtvrqvkveepsavkftqsg--ettidakepatedk-gikalakesakq 2772
QY 304 PPFASHPVHGYKLAPFASEVLAE-----AGRVAAELGH-----RLTHP-----QOFT 346
Db 2773 tpapaasvtgsrrrrprapresaqaiedlagfkpaaghteesmddtktkipcksspele 2832
QY 347 QLGSPRKEVESAIRDLEYHDELLSLKL-----PQQN 380
Db 2833 dtatskrtrpradkvevekeellavgklqtsgethtdkepvgeqgktafkqakpn 2892
QY 381 RDVAMIHHMGO---FGDKAATLERFKRNYARLSOSCKNRLVLENDVGVTVHDLPLVCE 437
Db 2893 vdaedvigsrrprapkeagpld--asfqlsgt----- 2927
QY 438 ELNIPWLDYHHNICFPDAHLRE-----GTLDISDP-KLOBRIANTWKRK 482
Db 2928 -----pghteelangaadsftsapktqdsqkplksirrvlrapk-- 2967
QY 483 GIKOKMHYSEPCDGAVTPRDRKKHPRVMT-LPPCPPDMDLMIKAKDKEQAVFELMRTFK 541
Db 2968 -----vepvgdvstddpvksqskntsllpplp-----fkgggkgdsvgtgkrlrc 3014
QY 542 LPEFEKINDVYDRDENRAPPVKKPKKKGGKRRKRTTDEAAEPEEVDTAADDVKDA 601
Db 3015 mpapeelieelpaskkqvraprgksspevvimkrlsrtasakrlpeae-einsndmktn 3073
QY 602 PRGPK---EVPEE 612
Db 3074 keehklqdsvpensk 3087

RESULT 13
AA14533
ID AA14533 standard; Protein; 2665 AA.
XX
AC AA14533;
XX
DT 12-OCT-2001 (first entry)
XX
DE Peptide #967 encoded by probe for measuring cervical gene expression.
KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX cervical cancer.
XX Homo sapiens.
OS
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.

```

```

XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX Claim 27; SEQ ID No 19359; 487pp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
XX (SENPs: see AA110068-AA128459). The present sequence is a peptide encoded
XX by one such probe. The SENPs are derived from human HeLa cells. The SENPs
XX can be used to produce a single exon microarray, which can be used for
XX measuring human gene expression in a sample derived from human cervical
XX epithelial cells. By measuring gene expression, the probes are therefore
XX useful in grading and/or staging of diseases of the cervix, notably
XX cervical cancer.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2665 AA:

```

```

Query Match 3.9%; Score 136; DB 22; Length 2665;
Best Local Similarity 18.8%; Pred. No. 0.034;
Matches 147; Conservative 95; Mismatches 296; Indels 242; Gaps 31;

```

```

QY 3 SRKSAAALDTPOSESTFSSTLD-----SSAPSP-----ARNLRRSGRNILQPSSEKD 51
Db 840 sreenwslfd-dsrfanfrnkdkkvdspripwmykxkkkirtldsegkmdkkeh 898
QY 52 RDHEKRSGBELAGRM-----GKDAN--GHCLRE-----GKEQEGVYKMAIEG 92
Db 899 keeegerqelfasrlfihssifedqskrlghlerkeedsdfisrgiygkqseanstds 958
QY 93 LAR-----MERLORATYKROKKQLEEDGIPVSVSRFPTAPVHHKSTNA-----EEREAKE 144
Db 959 iqepvvlfrsfmeltmrmqkkekd--qkpkveqkedtenhpktpesapenkdselkt 1016
QY 145 p-----VLKTHSKDVERAEIGVDDVVVMEPAATNIEPEDAQDAAE 186
Db 1017 ppsvgpsvtvtvlesapsalekttdgktveapl-vteektvepatsv---eeakpase 1071
QY 187 RGAARPPAVNSYLLP-----LPWKGRUGYACLTLYLRNAKPPITFSSRTCTMASI 235
Db 1072 papapveqlseqvdlppgadpkeaanmpagveegsgdppylidakpptpgasfsqaesn 1131
QY 236 VDHR-----HPLOFEDEPEHHLKKNPKDSKEPQDELGHKFVQELGLANARDIVKMLC 287
Db 1132 vdpdpdstpplskaqkseeaneapkaekpdadacpdangk----- 1173
QY 288 WNEKYGIRFLRLSSEMPFPASH-----PVHGYKLAPFASEVL 324
Db 1174 -----aaapesqpasedleivdpvaakdkkpnkskrsktpvqaaavsiavekpt 1224
QY 325 AEACRVAALCHRLTTHPGQFTQLGSPRKEVESAIRDLEYHDELLSLKLPPQONRDVAV 384
Db 1225 rkseridre-----klkrsnprgea-----qkllkmeaekirtas 1263
QY 385 MITHMGQFGDKAATLERFKRN-----YARLSQSKNRLVLENDVGVTVHDLPLVCEEL 439
Db 1264 knsaadiehpepslplsrttrrvsvyatmgdh-enr-----spvkepv 1307
QY 440 NIPW-----LDYHHHNICFPDAHLREGTLDISDPKLOERTANTWKRKGIKOKMHYSEPCD 495
Db 1308 eqprvtrkrlereleqaaaavpttprg-----rpbktrrr-adeeeneakepaetlkppe 1362
QY 496 GAVTPDRRR-----KHRPRV-MTLPCCPMDMLMIEAKD-----KQOAVF 534
Db 496 GAVTPDRRR-----KHRPRV-MTLPCCPMDMLMIEAKD-----KQOAVF 534

```


Db 1363 gwrspqktaaggpgqgkgnepkvdatrpeattevpgqigvkessmepkaaeaeags 1422
 QY 535 ELMRTFKLPGEKINDMVPYDRDDENRPAVPKPKKKG----- 574
 Db 1423 eqkrdkdagtdknppetavvevkkpape-knskskrgrsrnsrlavdksaslknvda 1481
 QY 575 -----GKRK-----RTTDEAAPEPEVDTAADVKDAPEGPKVEPE-EERAMGGP 618
 Db 1482 avspgaaqageregsvavspkespqkedglssqklsdpvdpdkekedvsasgp 1541
 RESULT 14
 AAM26950
 ID AAM26950 standard; Protein: 2665 AA.
 XX
 AC AAM26950;
 XX
 DT 17-OCT-2001 (first entry)
 XX
 DE Peptide #987 encoded by probe for measuring placental gene expression.
 XX
 KW Probe: microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00663.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 XX WPI; 2001-488897/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human placenta -
 XX
 PS Claim 27; SEQ ID No 27219; 654pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes (SENPs;
 CC see AA131315-AA157546). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for producing a microarray for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from human placenta. The probes are useful for antenatal diagnosis of
 CC human genetic disorders.
 XX
 SQ Sequence 2665 AA;

Query Match 3.9%; Score 136; DB 22; Length 2665;
 Best Local Similarity 18.8%; Pred. No. 0.034;
 Matches 147; Conservative 95; Mismatches 296; Indels 242; Gaps 31;
 QY 3 SRKSKAAALDTQSSSTSSSTLD-----SSAPSP-----ARNLRRGRNLLQPSSEKD 51
 Db 840 greenwslfdw-dsrfanfrnknkdkvdsapripwmykknkirtsgskmddkdh 898
 QY 52 ROHEKRSGEELAGRMW-----GKDNAN--GHCLURE-----GKPEEGYKWMATEG 92
 Db 899 keeqqelfasrlhssifedqskrlghlkerkeeddfisgriyqktsegansttds 958

QY 93 LAR-----MERRLQRATKQKQLEEDGIPVPSVVSFRFTAPYHHKSTNA-----BEREAKE 144
 Db 959 lqepvvlfhsrfmeltrmqkkekd--qkpkvekedqtenhpktpesapenkdselkt 1016
 QY 145 P-----VLKTHSKDVEREAIEIGVDVWVMEPAATNIIIEPEDAQDAAE 186
 Db 1017 ppsvgpsvtvltlesapsalekttgdkvtveapl-vteektvepatsv-----eakpase 1071
 QY 187 RCAARPPAVNSSLP-----LPWKGRLGYACUUNTYLRNAKPPIFSSRTRCMAASI 235
 Db 1072 papapveqlqevdlppgaadpkkaeamppagveegssgqppylakpptgpaasfsqaesn 1131
 QY 236 VDRH-----HPLQFDEPEHHLKNPKDKSKPEQDELGHKFVQELGLANARDIVKMLC 287
 Db 1132 vdepdstqplskpaqkseeanepkaekpdacadaepdangk----- 1173
 QY 288 WNEKYGIRFLKLSSEMFPPASH-----PVHGYKLAPFASBEVL 324
 Db 1174 -----aaapesqpasedievppvaakdkkpnkskrsktpvqaaavivekpvt 1224
 QY 325 AEAGRVAAELGHLRLTHPGQFTQLGSPRKEVVESAIRDELYHDELLSLKLPLEQONRDVA 384
 Db 1225 rkseridre-----klksnsprgea-----qkllkmeaekitrtas 1263
 QY 385 MTHMGOGFGDKAATLBERKRN-----VARLSQSKNRLVLENDVGVTVHDLPLVCEEL 439
 Db 1264 knsaadlehepsiplsfrtrrnrvsvyatmgdh-enr-----spvkepv 1307
 QY 440 NIPMV-----LDYHHHNICFDPAHLREGTLDISDPKLOBERTANTWKRKGIKOKMHYSEPCD 495
 Db 1308 eqprvtrkrlereiqeaaavptprg---rrpktrrr-adeeeeneakepaetllkpe 1362
 QY 496 GAVTPRRR-----KHRPRV-MTLPPCPPDMDLMEAKD-----KEQAVF 534
 Db 1363 gwrspqktaaggpgqgkgnepkvdatrpeattevpgqigvkessmepkaaeaeags 1422
 QY 535 ELMRTFKLPGEKINDMVPYDRDDENRPAVPKPKKKG----- 574
 Db 1423 eqkrdkdagtdknppetavvevkkpape-knskskrgrsrnsrlavdksaslknvda 1481
 QY 575 -----GKRK-----RTTDEAAPEPEVDTAADVKDAPEGPKVEPE-EERAMGGP 618
 Db 1482 avspgaaqageregsvavspkespqkedglssqklsdpvdpdkekedvsasgp 1541
 RESULT 15
 AAM02259
 ID AAM02259 standard; Protein: 2665 AA.
 XX
 AC AAM02259;
 XX
 DT 09-OCT-2001 (first entry)
 XX
 DE Peptide #941 encoded by probe for measuring human breast gene expression.
 XX
 KW Probe; human; breast disease; breast cancer; development disorder;
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
 XX
 OS Homo sapiens.
 XX
 PN WO200157270-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 29-JAN-2001; 2001WO-US00661.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast
XX
XX Claim 27; SEQ ID No 10999; 322pp; English.
XX
XX The present invention relates to novel single exon nucleic acid probes
CC (see AA10010-AAL10067). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for measuring human gene expression in
CC a human breast sample, where the probe hybridises at high stringency to a
CC nucleic acid expressed in the human breast. The probes are useful for
CC predicting, diagnosing, grading, staging, monitoring and prognosing
CC diseases of the human breast, particularly those diseases with polygenic
CC aetiology. The diseases include: breast cancer, disorders of development,
CC inflammatory diseases of the breast, fibrocystic changes, proliferative
CC breast disease and non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2665 AA;

Query Match 3.9%; Score 136; DB 22; Length 2665;
Best Local Similarity 18.8%; Pred. No. 0.034;
Matches 147; Conservative 95; Mismatches 296; Indels 242; Gaps 31;
QY 3 SRKSAALDTPQSESTFSSFLD---SSAPSP-----ARNLRRSRGNILQPSSEKD 51
DB 840 greenwflw-dsrfanfrnnkkekvdspripismymkxkirtsdsegmddkxkdh 898
QY 52 RDHEXRSBELAGRMW-----GKDAN--GCLRE-----GKEQEGYKMAIEG 92
DB 899 keeqerqelfasrfhsifeqdsqkriqlhlerkeedsdfisgriyqkqtseganstds 958
QY 93 LAR---MERRLORATKQKOLEEDGIPVSVVSRRFTAPYHHKSTNA---BEREAKE 144
DB 959 iqepvvlfhsrfmrlmqgkekd--qkpevekedtenhpktpesapenkdselkt 1016
QY 145 P-----VLKTHSKDVEREAEICVDVVRKMEPAATNIIPEDAQDAAE 186
DB 1017 ppsvppsvvtvlesapsalekttgdkttveapl-vteektvepatvs---eeakpase 1071
QY 187 RGAARPPAVNSSYLP-----LPWKGRGLYACLTLYLRNAKPPIFSSRTRCMASI 235
DB 1072 papapveqileqvdllppgdpkkaeammpagveegsgdqpypylakpbtptgafsfqaesn 1131
QY 236 VDHR-----HPLQFEDEPEHHLKNPKDSKEPQDELGHKFVQELGLANARDIVKMLC 287
DB 1132 vdpedstqplskpaqkseeaneapkaekpdatadaepdanq----- 1173
QY 288 WNEKYGIRFLRLSSEMFPPASH-----PVINGYKLAPFASEVL 324
DB 1174 -----aeapesqppasedleavdpvpaakkkpnkskrsktpvgaavsvivekpvt 1224
QY 325 AEAGRVAALGHRLTTHPGQFTQLGSPRKEVVEATRDLEYHDELLSLKLPEQONRDV 384
DB 1225 rksridre-----klkrinsprgea-----qklllekmaekitrtas 1263
QY 385 MIIHMGQGFQDKAATLERKRN-----YARLSQSKCNRLVLNDVDVGTVHDLPLVCEEL 439
DB 1264 knsaadlehepslplsrrrrnrvsvyatmgdh-enr-----spvkepv 1307
QY 440 NIPMV-----LDYHHHNCIFDPAHLREGTLDISDPKLOERIANTWKRRGKIKOKMHYSEPCD 495
DB 1308 eqprvtrklrelqeaavaavtptprg-----rppktrrr-adeeeeneakepaetlkppe 1362

QY 496 GAVTPRDRS-----KRRPV-MTLPPCPDMDLMEAKD-----KEQAVF 534
DB 1363 gwrsprrsqktaagggpggkkgknepkvdatrpeattevgpqigvkessmepkaaeags 1422
QY 535 ELMRTFKLFCFEKINDMVPYDRDDENRPPAPPVKAPKKKKG----- 574
DB 1423 eqkrdrkdsgrtdknppetapvevvekkpape-knskskrgrsrnsrlavdksasiknvda 1481
QY 575 -----GRRK-----RTTDEEAAAEPEEVDTAADDVKDAPEGPKVEPE-EERAMGCP 618
DB 1482 avspgaaagaderesgvvavspkespqkedglssqiksdqvpdpdkepekedvsasgp 1541

Search completed: January 15, 2002, 13:50:07
Job time: 296 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:52:23 ; Search time 74.94 Seconds
(without alignments)
666.807 Million cell updates/sec

Title: US-09-724-296-36

Perfect score: 3482

Sequence: 1 MPSRKSKAAALDTPQSEST.....REVKKGVPEVEDECFDG 656

Scoring table: BLOSUM62

Gap 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3472	99.7	656	2 S55262	UV-endonuclease -
2	1011.5	29.0	599	2 S71134	UV-endonuclease -
3	377	10.8	322	2 E83974	hypothetical prote
4	348.5	10.0	320	2 S55418	UV-endonuclease ho
5	261.5	7.5	326	2 C75350	probable UV damage
6	160.5	4.6	1280	2 T00365	hypothetical prote
7	155.5	4.5	5327	2 T13564	microtubule-associ
8	146	4.2	796	2 S56231	hypothetical prote
9	141.5	4.1	1110	2 I51116	NF-180 - sea lamp
10	140	4.0	727	2 T17292	hypothetical prote
11	139.5	4.0	1510	2 T33100	hypothetical prote
12	139	4.0	1815	2 C81169	IgA-specific metal
13	138	4.0	1547	2 T28657	blackjack protein,
14	137	3.9	1560	2 T42727	proliferation pote
15	137	3.9	2897	2 B48666	cell proliferation
16	137	3.9	3256	2 A48666	cell proliferation
17	136.5	3.9	845	2 A48666	neurofilament trip
18	136.5	3.9	849	2 S00030	neurofilament trip
19	135.5	3.9	3942	2 T42730	Bassoon protein -
20	135.5	3.9	971	2 T24866	hypothetical prote
21	135	3.9	793	1 JH0628	caldesmon - human
22	134	3.8	2663	1 S28261	centromere protein
23	133.5	3.8	980	2 E71606	hypothetical prote
24	132.5	3.8	915	2 S54485	CES1 protein - yea
25	132.5	3.8	6642	2 T29757	protein UNC-89 - C
26	132	3.8	734	2 F85073	hypothetical prote
27	131.5	3.8	1222	2 T24420	hypothetical prote
28	131	3.8	606	2 A43427	neurofilament trip
29	128.5	3.7	856	2 T16543	hypothetical prote

30	128.5	3.7	7962	2 I38346	elastic titin - hu
31	128	3.7	1020	1 QFHHH	neurofilament trip
32	128	3.7	1805	2 A34736	nestin - rat
33	128	3.7	3488	2 T34418	hypothetical prote
34	128	3.7	4151	2 T13734	groovin gene prote
35	127.5	3.7	2453	2 S60254	nuclear receptor c
36	126.5	3.6	720	2 T26819	hypothetical prote
37	126.5	3.6	2416	2 T13825	adenomatous polyo
38	126	3.6	522	2 C96608	hypothetical prote
39	126	3.6	1773	2 A81937	IgA-specific metal
40	125.5	3.6	786	2 C86406	hypothetical prote
41	125.5	3.6	1400	2 T52359	hypothetical prote
42	125.5	3.6	2441	2 D71623	erythrocyte membra
43	124.5	3.6	1829	2 T26135	hypothetical prote
44	124.5	3.6	4377	2 A55575	ankyrin 3, long sp
45	124	3.6	1072	1 A37221	neurofilament trip

ALIGNMENTS

RESULT 1

S55262

UV-endonuclease - Neurospora crassa

C:Species: Neurospora crassa

C>Date: 10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 20-Jun-2000

C:Accession: S55262

R:Yajima, H.; Takao, M.; Yasuhira, S.; Zhao, J.H.; Ishii, C.; Inoue, H.; Yasui, A.

EMBO J. 14, 2393-2399, 1995

A:Title: A eukaryotic gene encoding an endonuclease that specifically repairs DNA dam

A:Reference number: S55262; MUID:95292980

A:Accession: S55262

A>Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-656 <YAJ>

A:Cross-references: EMBL:D11392; NID:q1526560; PID:g927215

C:Genetics:

A:Gene: uvel

Query Match	99.7%	Score	3472	DB 2:	Length	656;			
Best Local Similarity	99.7%	Pred. No.	2.2e-208;						
Matches	654;	Conservative	1;	Mismatches	1;	Indels	0;	Gaps	0;
Oy	1	MPSRKSKAAALDTPQSESTFSSTLSSAPARNLRSGRNILQPSSEKDRDHRKRSGE	60						
Db	1	MPSRKSKAAALDTPQSESTFSSTLSSAPARNLRSGRNILQPSSEKDRDHRKRSGE	60						
Oy	61	ELAGRMGKDANGCHLREGKQEEGVKMAIEGLARMEERRIQRATKROKKOLEEDGTPVPS	120						
Db	61	ELAGRMGKDANGCHLREGKQEEGVKMAIEGLARMEERRIQRATKROKKOLEEDGTPVPS	120						
Oy	121	VVSFPPTAPYHHKSTNAEREAKPEVLKTHSKDVERAEIGVDDVVVKMPEPATNIIIEPED	180						
Db	121	VVSFPPTAPYHHKSTNAEREAKPEVLKTHSKDVERAEIGVDDVVVKMPEPATNIIIEPED	180						
Oy	181	AQDAEERGAARPPAVNSSYLLPLPKWGLGYACINTYLRNKKPPIESSRTRCMASIVDHRH	240						
Db	181	AQDAEERGAARPPAVNSSYLLPLPKWGLGYACINTYLRNKKPPIESSRTRCMASIVDHRH	240						
Oy	241	PLQFEDEPEHLLKNPKDKSEPODELCHKFVEQELGLANARDIVKMLCWNEKYGIIFRLRLS	300						
Db	241	PLQFEDEPEHLLKNPKDKSEPODELCHKFVEQELGLANARDIVKMLCWNEKYGIIFRLRLS	300						
Oy	301	SEMPFFASHPVHGYKLAPFASEVLAEGRAVAAELGHLRTHHPQOFTQLGSPRKEVVEISAI	360						
Db	301	SEMPFFASHPVHGYKLAPFASEVLAEGRAVAAELGHLRTHHPQOFTQLGSPRKEVVEISAI	360						
Oy	361	RDLEYHDELLSLKLPQONRODAVMIHMGQFGDKAAATLERFKRNYARLSQSCKNRVL	420						
Db	361	RDLEYHDELLSLKLPQONRODAVMIHMGQFGDKAAATLERFKRNYARLSQSCKNRVL	420						
Oy	421	ENDDVGWTVHDLPLVCEELNIPMVLVDYHHHNTCFDPAHLREGTLDISDPKQLOERIANTWK	480						

Db 421 ENDDVGVTHDLPVCELNIPVLDYHHHNCFCFAHLREGTLDISDPKLOBRANTWK 480
 QY 481 RKGIKQKMHYSEPCDGAATPDRKRHRPRVMTLPCCPDMDLMEAKDKEQAVFELMRTF 540
 Db 481 RKGIKQKMHYSEPCDGAATPDRKRHRPRVMTLPCCPDMDLMEAKDKEQAVFELMRTF 540
 QY 541 KLPGFEKINDVVPYDRDENDRPPVVPKAPKKKGKRRKTTDEEAPEPEVDVTAADDVKD 600
 Db 541 KLPGFEKINDVVPYDRDENDRPPVVPKAPKKKGKRRKTTDEEAPEPEVDVTAADDVKD 600
 QY 601 APEGKPEPEERAMGGPNRYWPLGCEEWLKPCKREVKKGVPEVEDEGEFBDG 656
 Db 601 APEGKPEPEERAMGGPNRYWPLGCEEWLKPCKREVKKGVPEVEDEGEFBDG 656

RESULT 2
 S71134
 UV-endonuclease - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 20-Jun-2000
 C:Accession: S71134; T39815
 R:Takao, M.; Yonemasu, R.; Yamamoto, K.; Yasui, A.
 Nucleic Acids Res. 24: 1267-1271, 1996
 A:Title: Characterization of a UV endonuclease gene from the fission yeast Schizosaccharomyces pombe
 A:Reference number: S71134; MUID:96188860
 A:Accession: S71134
 A:Status: preliminary; nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-599 <TAK>
 A:Cross-references: EMBL:D78571; NID:gi399000; PIDN:BAAL1415.1; PID:gi399001
 A:Experimental source: strain S972
 R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Lucas, M.; Gaillardin, C.
 submitted to the EMBL Data Library, June 1998
 A:Reference number: Z21881
 A:Accession: T39815
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-599 <LYN>
 A:Cross-references: EMBL:AL023859; PIDN:CAA19577.1; GSPDB:GN00067; SPDB:SPBC19C7.09c
 A:Experimental source: strain 972h; cosmid cl9C7
 C:Genetics:
 A:Gene: UVDE; SPBC19C7.09c
 A:Map position: 2

Query Match 29.0%; Score 1011.5; DB 2; Length 599;
 Best Local Similarity 36.6%; Pred. No. 1.9e-55;
 Matches 241; Conservative 88; Mismatches 149; Indels 181; Gaps 18;

QY 3 SRKSAALDTPQSESTFSTLDDSSAPSPARNLRSGRNILQPSSEKDRDHRKRSGBEL 62
 Db 110 ARSKKKATDSVSKIDESVASYDSST-----HLRRSSR-----SKKPVNYSSESE- 157
 QY 63 AGRMGKDANGCHLUREGKEGEGVMAIEGLARMERLORATKROKKOLEED----- 114
 Db 158 -----SEQISKATKKVKKKEEEVEVEVDEK 184
 QY 115 GIPVSVVSRF-PTAPYHHKSTNAEAEKPEVLKTHSKDVERAEIGVDDVVKMEPAAT 173
 Db 185 SLKNSSSDEEPVVPPEOLETTPISKRRS-----RSSAKNLEKSTMNLDHHA----- 232
 QY 174 NIIEPEDAQAARGAARPPAVNSYLLPDKWKGRLGYACINTYLRNAKPPIPSSRTCRMA 233
 Db 233 -----PREMFDCDK-----PIWRGRLGYACINTYLRNAKPPIPSSRTCRIT 275
 QY 234 STVDHRHPLQFEDEPEHHLKNKPKSEPODELCHKFQVQELGLANARDIVKMLCWNEYK 293
 Db 276 TI-----QROGLESVKQLGTQNVLDLIKLVENHNEG 307

QY 294 IRFLRLSEMFPFASHPVHGYKLAFFASEVLAEGRAVAELGRLTHTHPCQFTQLGSPRK 353
 Db 308 IHFMVSSDLFPFASHAKYGYTL-EFAQSHLEEVGKLANRYNHLTMHPQOYQTIASPRE 366

QY 354 EVVESAIRDLEYHDELLSLKLPQGNRDVAMLIHMGOGFQDKAATLIERKRNARLSQS 413
 Db 367 VVDSAIRDAYHDEILSRMKLNEQUNQKRAVLIIHLCGTFEGKKEKTLDRKRNQRLSDS 426
 QY 414 CKNR[LVLENDVGVTVHDLPLVCEELNIPVLDYHHHNCFCFAHLREGTLDISDPKLOE 473
 Db 427 VKARLVLENDVGVTVHDLPLVCEELNIPVLDYHHHNCFCFAHLREGTLDISDPKLOE 482
 QY 474 RIANTWRKGIQKMHYSEPCD-GAVTPDRKRHRPRVMTLPCCPDMDLMEAKDKEQ 532
 Db 483 TIRETWIRKGIQKMHYSEPCD-GAVTPDRKRHRPRVMTLPCCPDMDLMEAKDKEQ 542
 QY 533 VFELMRTFKLPGFEKINDVVPYDRDENDRPPVVPKAPKKKGKRRKTTDEEAPEPEVD 592
 Db 543 VFELMRTFKLPGFEKINDVVPYDRDENDRPPVVPKAPKKKGKRRKTTDEEAPEPEVD 592

RESULT 3
 E83974
 hypothetical protein BH597 [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
 C:Accession: E83974
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Fujii, F.; H
 Nucleic Acids Res. 28: 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
 A:Reference number: A83650; MUID:20263314
 A:Accession: E83974
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-322 <STO>
 A:Cross-references: GB:AP001516; GB:BA000004; NID:gi10175192; PIDN:BAB06316.1; GSPDB:G
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH2597

Query Match 10.8%; Score 377; DB 2; Length 322;
 Best Local Similarity 29.1%; Pred. No. 2.4e-16;
 Matches 105; Conservative 61; Mismatches 127; Indels 68; Gaps 10;

QY 207 RLGAYACINTYLRNAKPPIPSSRTCRMASIVDHRHPLQFEDEPEHHLKNKPKSKPEQDEL 266
 Db 4 QFGYVAMSMELANASP-----SKTMTAT-----QFEKTEDH-----EA 36
 QY 267 GHKFOVQELGLANARDIVKMLCWNEYKIGIRFLRLSEMFPFASHPV-HGYKLAFFASEVLA 325
 Db 37 GLKRLERIAKTNLHNCRLKLLKUNLAQISFRLSKSLVPLVNHPLTEGKVELAIAEELQ 96
 QY 326 EAGKVAELGRLTHTHPCQFTQLGSPKREVESAIRDLEYHDELLSLKLPQGNRDVAM 385
 Db 97 AVGEFASHEQMRIDFPHDFHVLNLSAKETIRRSQTLIDYHYKLLKLGMEIDPRHR----C 152
 QY 386 LIHMGOGFQDKAATLIERKRNARLSQSCKNRILVLENDVGVTVHDLPLVCEELNIPWL 445
 Db 153 VLVHGGKRAKVEACIQFQFENTASIPKSLLSMLMLENDOKSYTIDVLYLGEKLAIPVL 212
 QY 446 DYHHHNCFCFAHLREGTLDISDPKLOERITANTYLRNAKPPIPSSRTCRMA 501
 Db 213 DIHHHDLHRSKSLQE-----TWQIRVATWEDSPVPIHLSPLSGEDDPRHHDY 263
 QY 502 ---DRKRHRPRVMTLPCCPPD---MDLMEAKDKEQAVFELMRTFKLPGFEKINDVVPY 554
 Db 264 INADR-----FIAPLHEITGADAVHDLHVMLEAKKDLALFQLMK-----DLAEY 307

QY 555 D 555
 Db 308 D 308

[illegible]

[illegible]

Db 1326 TSLATQKOSSTTGKVP 1342

RESULT 14

T42727

proliferation potential-related protein - mouse

C:Species: Mus musculus (house mouse)

C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 02-Sep-2000

C:Accession: T42727

R:Witte, M.M.; Scott, R.E.

submitted to the EMBL Data Library, November 1998

A:Reference number: 222246

A:Accession: T42727

A:Molecule type: mRNA

A:Residues: 1-1560 <WIT>

A:Cross-references: EMBL:U83913; NID:g3858884; PID:g3858885; PIDN:AAC72432.1

A:Experimental source: strain Balb/C

C:Genetics:

A:Gene: P2P-R

C:Function:

A:Description: involved in hnRNP association and Rb1 binding

C:Superfamily: RING finger homology

F;57-107/Domain: RING finger homology <RRN>

Query Match 3.9%; Score 137; DB 2; Length 1560;

Best Local Similarity 19.3%; Pred. No. 1.6;

Matches 157; Conservative 100; Mismatches 326; Indels 232; Gaps 33;

QY 2 PSRSKAAALDTP-QSESSTFSSTLSSAPARNLRSGRNILQPSSEKDRDHEKRSGE 60

DB 749 PSREDATPVREDEMDAESITFKVSVD-----KDKREKDKPKVSKDKTKRSKSDGS 797

QY 61 ELAGRMGMKDANGHCLREGQEKGVEKMAIEGLARMERLQORATKROKOLEBDGIPVPS 120

DB 798 ATAKK-----DNVLPKSGQKQVGDREKSPRSEPLPKA-----KEEATKIDS 842

QY 121 V-----VSRPTAPYHHKSTNAERAK-----EPVLKTHSKDVERAEIGVDDV 165

DB 843 VKPSSSOKDEKVTGTGRKA--HKSAAKEHQEAKPAKDEKVKKDCSKDKSEKPSKDEK 900

QY 166 VKMPEAATNII-----EPEDAQAARGAARPPAVNSYLLPLPWKGLGYACLN 214

DB 901 AK-KPEKNKLLDSKGEKRRKTEESKVDKDFESSMKISKVEGTEIVKSPKMEGDVE 959

QY 215 TYLRNAKPPIFSSRTCRMASIVDHRHPLOFEDEPEHHLKNKPKDS--KEPQDEL---GHK 269

DB 960 KLERTPEKDKIASSTTPPAKKIKLNR-----ETGRKKIGNAENASTTKPESEKLESTSSK 1012

QY 270 FVQELGLANA-----EPEDAQAARGAARPPAVNSYLLPLPWKGLGYACLN 214

DB 1013 IKQEVKVKAKRVKAGSESSLTVDYTSSTGSGPVRKSEKTDTRIVIKTM--EE 1069

QY 292 YGIRFLKLSSEMPFASHPVHGYKLPAPASE-----VLADAGRAVAALGHLRTHP- 342

DB 1070 YNDNTAPAEVDVLIHVPQSKDKDDDFESEEDVKTTPQIQSGVKPSSII-KNVTTKPS 1128

QY 343 --GOFTOLGSPKEVVE--SAIRLDYHDELILLSLKLPQEQNRDAVMIIMG----- 390

DB 1129 ATAKYTEKESEQEKLOKLPKEASHLMQH-ELRSKSGSASSKGRADREHSGSKDNP 1187

QY 391 -----GQFGKAKATLERFKR--NYARLSQSKNRLVLENDV--GWTVHDLILPVCEELNI 441

DB 1188 DKRSGAQDPKESTVDRLSQGHFKTLSQSKETRTSEKHESVRGSSNKDETPGRDK--- 1244

QY 442 PMVLDYHHNHCIDPAHLRGTI-----DISDPKLOERANTWKKIGIKQKMH 489

DB 1245 --KVYDSRYSKSKRDERGELARKDPPRKESLSGOKSKLREERDLPPKGAESRKS 1302

QY 490 YSEPCDAGVTPPRDRKH-----RPRVMTLP----- 514

DB 1303 NSSP-----PRDKKPHDKAPYETKRPCEETKPVDKISKEREKHAAREAKDSSGK 1356

QY 515 -PC---PPLMDLMIKAKDEQAVFELMKTFKLP-----GFEKINDMVPYDRDD 558

DB 1357 LFCILNPPULPM-----EKELAVGQVEKSAVKPKPOLSHSSRLSSDLTRETDEAAFEPTY 1411

QY 559 ENRPAPPVKAPKKK-----GKKRRRTTDEAAEPVEEDTAAD---DVKDAPEGPKVEPE 610

DB 1412 NESDSSESNVKKEEAVASISKDLKEKTEKAKESLTVATASQPCADRQSQSSPSVSPS 1471

QY 611 EERAMGGPYNRVY----WPLGCEEWLKKPKREVKK 641

DB 1472 RSHSPSGQTRSHSSASSAGSQDSKKKKKKKKKK 1506

RESULT 15

B48666

cell proliferation antigen Ki-67, short form - human

C:Species: Homo sapiens (man)

C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 07-May-1999

C:Accession: B48666

R:Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Ge

A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, u

ins.

A:Reference number: A48666; MUID:94043435

A:Accession: B48666

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-2897 <SCH>

A:Cross-references: EMBL:X65551

C:Superfamily: kinase interaction domain homology

C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat

F;29-91/Domain: kinase interaction domain homology <KIH>

Query Match 3.9%; Score 137; DB 2; Length 2897;

Best Local Similarity 19.9%; Pred. No. 3.6;

Matches 146; Conservative 92; Mismatches 296; Indels 200; Gaps 30;

QY 5 KSKAALD-----TP--QSESSTFSSTLSSAPSPA-----RNLRSGRNILQPS 48

DB 2069 KEKAEALDVGFKELFOTPGHTEESMTDDKITEVSKSPQPESEFKTSRSKQRLKPLV 2128

QY 49 EKDRDHEKRSGEELAGRMGKDANGHCLREG-----KEQESGVKMAIEGLARME-RLQR 102

DB 2129 KVMKEELAVSKLT-RYSGETTQTHTEPTGDSKSIKAFKESPKOILDPAASVTGSRRL 2187

QY 103 ATKROKOLEEDGIPVPSVWSRFPATPYHHKSTNAEAREKEPV----- 146

DB 2188 KTRKEKARALDVLDFKELFS---APGHTESMTIDKNTKIPCKSPPELTDTATSKR 2243

QY 147 -----LKTHSKDVERAEI-GVDDVVKMEPA-----TNII-----EPEDAQA 186

DB 2244 CKTPRKREKVEELSAVERLTQTSQSTHTHKEPASGDGKVLKORAKKKNPNVEEPPS 2303

QY 187 RGAARPPAVNSYLLPLPWKGLGYACLNLYLRNAKPPILFSSRTCRMASIVDHRHPLOFED 246

DB 2304 RRRPRAPKEAQ---PLEDLAGTELSETSGHTQESLTACKATK---IPCESPLEWD 2356

QY 247 ---EPEHILKNKPUSKEPEQDELGHKFWOELGLANARDIVKMLCWNKYGIRFLRLSEM 303

DB 2457 TJATSKRLHRTVQVQVQKEEPSAVKFTQTSQ---ETTDADKEPAGEDK-GIKALKESAKQ 2413

QY 304 PPFASHPVIGYKLPAPFASEVLAE-----AGRVAAELGH-----RLTTHP-----GOFT 346

DB 2414 TPAAASVTGSRRRPRAPRESQAQIEDLAGKDPAAAGHTESMTDDKTKIPCKSSPELE 2473

QY 347 QIGSPRKEVESAIKRLDYHDELILLSKL-----PEQON 380

DB 2474 DTATSSKRRPRTRAKQVEKVELLAVGLKLTQTSGETTHTDKEPVGEGKTKAFKQPAKRN 2533

QY 381 RDVMTIIMGQ---FGDKAATLERFKFNARLSQSKNRLVLENDLVGTVTHLLPVCE 437

```
Db 2534 VDAEDVIGSRQPRAPKEKAQPLEDL-ASFQELSQT----- 2568
Qy 438 ELNIPWVLDYHHHNICFDPAPHLRE-----GTLDISDP-KLOERIANTWKRR 482
Db 2569 -----PGHTEELANGAADSFTSAPKQTPDSGKPLKISRRLRAPK-- 2608
Qy 483 GIKQKMHYSEPCDGAVTPRDRRKHPRVMT-LPPCPPDMDLMEAKDKQAVFELMRTEK 541
Db 2609 -----VEPVGDVSTRDPVKSQSKSNTSLPPLP-----FKRGGKDGSVTGTKRLRC 2655
Qy 542 LPGFEKINDRVYDRDDENRAPPPVAPKAPKKKGKRRKTTDEEAAPPEEVDTAADDVKDA 601
Db 2656 MPAPPEIVEELPASKKQORVAPRARGKSSEPVWIMKRSRLRTSAKRIPAE-ELNSNDMKTN 2714
Qy 602 PEGPK-----EVPPEE 612
Db 2715 KEEHKLQDSVPENK 2728
```

Search completed: January 15, 2002, 13:52:31
Job time: 400 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:49:55 ; Search time 120.01 seconds
(without alignments)
369.718 Million cell updates/sec

Title: US-09-724-296-2_COPY_230_828

Perfect score: 3112

Sequence: 1 MLRLKRNQIKSRIVFTIL.....KRLTARKRSRKEVEDEK 599

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_1101.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3112	100.0	828	21 AAY44498	GST signal peptide
2	1979	63.6	371	21 AAY44499	S. pombe delta228-
3	1979	63.6	600	21 AAY44500	GST signal peptide
4	1011.5	32.5	656	17 AA001618	Neurospora crassa
5	961.5	30.9	626	21 AAY44502	N. crassa delta228
6	961.5	30.9	626	21 AAY44503	B. subtilis delta2
7	230.5	7.4	294	21 AAY44505	D. radiodurans del
8	147.5	4.7	782	11 AAR06991	Polypeptide antigen
9	147.5	4.5	612	22 AAB95546	Human protein sequ
10	140.5	4.5	733	19 AAW60785	Deletion mutant de
11	140.5	4.5	2441	21 AAB18161	Plasmodium falcipa

12	136.5	4.4	1639	19 AAW54145	P. falciparum synt
13	135	4.3	1341	21 AAY85657	Human Acinus L pro
14	134.5	4.3	753	19 AAW60788	Deletion mutant de
15	134.5	4.3	753	19 AAW60787	Deletion mutant de
16	134.5	4.3	1654	6 AAP50777	Sequence of the P1
17	134	4.3	2688	22 AAM40883	Human polyprotein
18	133.5	4.3	1588	15 AAR46605	Malarial PfEMP3 ep
19	133.5	4.3	1663	15 AAR46608	Plasmodium falcipa
20	133	4.3	677	20 AAW85608	Secreted protein c
21	133	4.3	677	20 AAW85608	Human myosin heavy
22	132.5	4.3	2663	22 AAG64560	Human polyprotein
23	132	4.2	414	17 AAM39097	Saccharomyces cere
24	132	4.2	414	22 AAB10529	Amino acid sequenc
25	131.5	4.2	922	22 AAB30809	Human protein sequ
26	130.5	4.2	1197	22 AAB93218	Human protein sequ
27	130.5	4.2	1540	20 AAB93006	Human htlp2 protei
28	130.5	4.2	1674	19 AAY07734	Human BAZ1-alpha p
29	130	4.2	377	22 AAW81169	Peptide #4635 enco
30	130	4.2	377	22 AAM18201	Peptide #4733 enco
31	130	4.2	377	22 AAM30696	Peptide #4500 enco
32	130	4.2	615	21 AAG18697	Arabidopsis thalia
33	130	4.2	615	21 AAG50810	Arabidopsis thalia
34	130	4.2	633	21 AAG18696	Arabidopsis thalia
35	130	4.2	633	21 AAG50809	Arabidopsis thalia
36	130	4.2	662	21 AAG18695	Arabidopsis thalia
37	130	4.2	662	21 AAG50808	Arabidopsis thalia
38	130	4.2	1135	21 AAY84460	Amino acid sequenc
39	130	4.2	1257	22 AAB66926	RBP1 protein. Hom
40	129.5	4.2	431	20 AAY49137	Amino acid sequenc
41	129.5	4.2	489	22 AAB92515	Human protein sequ
42	129	4.1	1400	22 AAB83348	AAP-2 protein sequ
43	128.5	4.1	769	19 AAW60795	Deletion mutant de
44	128.5	4.1	769	19 AAW60796	Deletion mutant de
45	127.5	4.1	769	19 AAW60793	Deletion mutant de

ALIGNMENTS

RESULT 1
AAY44498
ID AAY44498 standard; Protein: 828 AA.

AC AAY44498:

DT 27-MAR-2000 (first entry)

DE GST signal peptide and S. pombe UVDE fusion protein.

KW GST signal peptide; ultraviolet damage endonuclease: UVDE; Uvelp;

KW UV irradiation; DNA damage; UV radiation damage; fusion protein;

KW skin cancer; glutathione-S-transferase.

XX Schizosaccharomyces pombe.

OS Synthetic.

XX

PH Key Location/Qualifiers

FT Peptide /label= GST_signal_peptide

FT Protein /note= "S. pombe UVDE"

FT MISC-difference 11

FT MISC-difference 85 /note= "Encoded by aaaa"

FT MISC-difference 147 /note= "Encoded by gtt"

FT MISC-difference 147 /note= "Encoded by gtt"

XX

PN

XX

PD

XX

XX

PF

08-JUN-1999; 99WO-US12910.

Db 181 kketldfrknygrisdsvkariwlvnddsvsqdlplcqlnlpvlwvhhnviyg 240
 Qy 469 TLREGSLDMLPIPTIRETWRTKGIQKOHYESADPTAISGMKRAHSDRVDFPPCDP 528
 Db 241 tlregslidmpliptiretwrtkgitqkghysesadptaisgmkrhahsdvdfppcdp 300
 Qy 529 TMDLMIEAKEQAVFLCRRYELQNPCCPLEIMGPEYDQTRDGYPPGAEKRLTARKRR 588
 Db 301 tmdlmieakeqavfclrryelnppcpleimgpeydtqrdgyppgaekrltarkrr 360
 Qy 589 SRKEEVEDEK 599
 Db 361 srkeeeveedek 371

RESULT 3
 AAY44500
 ID AAY44500 standard; Protein: 600 AA.

XX AAY44500;

XX 27-MAR-2000 (first entry)

XX GST signal peptide and delta228 S. pombe UVDE fusion protein.

XX GST signal peptide; glutathione-S-transferase signal peptide;

KW ultraviolet damage endonuclease; UVDE; delta228-UVDE;

KW repair-deficient E. coli strain; UV irradiation; DNA damage;

KW UV radiation damage; photoproduct; abasic site; platinum diaduct;

KW mismatched nucleotide pairing; nucleotide alkylation;

KW fusion protein; skin cancer.

XX Schizosaccharomyces pombe.

OS Synthetic.

XX Key Location/Qualifiers

XX Misc-difference 11 /note= "Encoded by aaaa"

FT Misc-difference 85 /note= "Encoded by gtt"

FT Misc-difference 147 /note= "Encoded by gtt"

FT Misc-difference 147 /note= "Encoded by gtt"

XX WO9963828-A1..

XX 16-DEC-1999.

XX 08-JUN-1999; 99WO-US12910.

XX 08-JUN-1998; 98US-0088521.

XX 18-MAY-1999; 99US-0134752.

XX (UYEM-) UNIV EMORY.

XX Doetsch PW, Kaur B, Avery AM;

XX WPI: 2000-116417/10.

XX N-PSDB: AAZ29860.

XX A new truncated ultraviolet damage endonuclease for treatment of skin

XX cancers -

XX Disclosure; Page 56; 133pp; English.

XX The present sequence is a fusion protein comprising the GST signal

XX peptide and delta228 truncated ultraviolet damage endonuclease (UVDE)

XX from S. pombe. Delta228-UVDE is a 288 amino acid deletion of the

XX N-terminal of the S. pombe uvrl+ gene product. This is expressed in frame

XX with a GST leader sequence. Stable endonuclease fragments can be produced

XX for cleaving a double-stranded DNA molecule that has a distorted

XX structure resulting from UV radiation damage, a photoproduct, an abasic

XX site, mismatched nucleotide pairing, platinum diaduct, an intercalated

CC molecule or alkylation of a nucleotide. UVel can be used in compositions
 CC for internal or topical application and as a therapeutic agent for skin
 CC cancers.

XX Sequence 600 AA;

Query Match 63.6%; Score 1979; DB 21; Length 600;

Best Local Similarity 100.0%; Pred. No. 3.8e-153;

Matches 371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 229 DDHAPREMFDCLDKPIPWGRGLGYACLNITLRSMKERVFCSTCRITTIQRDGLSEVKOL 288

Db 230 ddhapremfddcldkplpwgrglgyaclnitlrsmkervfcsrtcrittitqrdglesvqkl 289

Qy 289 GTQNVLDLILKLVENHNFCIHFMVRVSSDLFPFASHAKYCYTLEFAQSHLEEVCKLANKYN 348

Db 290 gtqnvldlilkvewnhnfcghfmrvssdlfpfashakgytlfafaqshleevgklankyn 349

Qy 349 HRLTMHPGQYTOITASPREVVVDSAIRDLAYHDEILSRMKLNQKNDAVLIILHGGTFEG 408

Db 350 hrltmhpgygtiasprevvvdsairdlayhdeilsrmlkneqlnkdavliihlggtfeg 409

Qy 409 KKETLDRFRKNYQRLSDSVKARLVLENDVSVSQDLPLCQELNIPVLVDWHHNNIVPG 468

Db 410 kketldfrknygrisdsvkarlvleendvsvsqdlplcqlnlpvlvdwhhnnivpg 469

Qy 469 TLREGSLDMLPIPTIRETWRTKGIQKOHYESADPTAISGMKRAHSDRVDFPPCDP 528

Db 470 tlregslidmpliptiretwrtkgitqkghysesadptaisgmkrhahsdvdfppcdp 529

Qy 529 TMDLMIEAKEQAVFLCRRYELQNPCCPLEIMGPEYDQTRDGYPPGAEKRLTARKRR 588

Db 530 tmdlmieakeqavfclrryelnppcpleimgpeydtqrdgyppgaekrltarkrr 589

Qy 589 SRKEEVEDEK 599

Db 590 srkeeeveedek 600

RESULT 4

AAW01618

ID AAW01618 standard; Protein: 656 AA.

XX AAW01618;

XX 20-AUG-1997 (first entry)

XX Neurospora crassa DNA repair enzyme.

XX DNA repair enzyme; cyclobutane-type dimer; (6-4)-bound product;

KW exposure; UV light; ultraviolet; suntan cream; prevention; treatment;

KW skin cancer.

XX Neurospora crassa.

XX JP08266275-A.

XX 15-OCT-1996.

XX 29-MAR-1995; 95JP-0094137.

XX 29-MAR-1995; 95JP-0094137.

XX (SHIS) SHISEIDO CO LTD.

XX WPI: 1996-512663/51.

XX N-PSDB: AAT58286.

XX DNA repair enzyme and related DNA - specifically recognises

PT cyclobutane-type dimer and (6-4)-bound product induced by exposure

XX to UV light

PS Disclosure; Page 10-13; 18pp; Japanese.

XX The present sequence shows a DNA repair enzyme which specifically
CC recognises a cyclobutane-type dimer and (6-4)-bound product induced,
CC respectively from TT and TC sequences in DNA, by exposure to UV light.
CC The enzyme cleaves these sequences at the 5'-side. The DNA repair enzyme
CC can be used in, e.g. suntan cream for prevention and treatment of skin
CC cancer. Conventional endonuclease acts on either the cyclobutane-type
CC dimer or the (6-4)-bound product in damaged DNA, while the enzyme
CC of the invention can act on both of them.

XX Sequence 656 AA;

Query Match 32.5%; Score 1011.5; DB 17; Length 656;

Best Local Similarity 36.6%; Pred. No. 5.2e-74;

Matches 241; Conservative 88; Mismatches 149; Indels 181; Gaps 18;

QY 110 ARSSKKKATDSVSDKIDSVASYDSST-----HLRRSSR-----SKPVPVNNSSSESE- 157

Db 3 srkskaaaidtpqsesstfsstldssapsparnlrrsgnllqpsssekdrdhekrsgeel 62

QY 158 -----SSEQISKATKKVKKEEEYVEEVDEK 184

Db 63 agrmmgkdanghclregkeegvkmaleglarmerriqratkrgkqgleed----- 114

QY 185 SLKNSSSDEFPVPVPEOLETPIKRRRS-----RSSAKNLEKESTMNLDHHA----- 232

Db 115 gipvpvsvrft-ptapyhhkstnaeereakepvltkshdkdveraeagvddvkmepaat 173

QY 233 -----PREMFDCLDK-----PIPWGRGLGYACLTILRSKMERVFCSCRTCRIT 275

Db 174 nilepedaqdaaergaarppavpnssylplpwkgrlgyacintylrnsakpplfsrtcma 233

QY 276 TI-----QRDGLSVKQLGTQNVLDLKLVEVNNHFG 307

Db 234 sivdhrhplqfedephehlknkpdkskepqdelghkfvgelglanardivkmlewnesqy 293

QY 308 IHFMVSSDLFFPFASHAKYGYTL-EFAQSHLEEVGKLANKYNHRLTMHPGQYTOIASPRE 366

Db 294 lrflrlsemfpfashpvhgylapfasevlaeagrvaaelghrltthpgqftqlgsprk 353

QY 367 VVVDASIRDLAYHDLIRSMKLNQNKDAVLIILHGGTFEGKKEKTLDRFRKNYORLSDS 426

Db 354 evvesairdleyhdelisllkpeqndravmihmgqgfdkaatlfrkryarlsqgs 413

QY 427 VKARVLVNDVSVSQDLLPLCOELNPLVLVDWHHNNIV--POTLRGSLDLM--PLIP 482

Db 414 cknrlvlendvgtvchdlpvceelnlpvldyhhnncfcpahlregtldisdpklqe 473

QY 483 TIRETWRTKGTOKOHYSSESADPTAISGMKBRASDRVDFPPCDPTMDLIMEAKEKEQA 542

Db 474 riantcwkrtgikqkmhysepcd-gavtprhrkrprvmtlppcpdmdlmieakkega 532

QY 543 VFELCRRYEL-----QNPPCP-----LEIMGPE-- 565

Db 533 vfelmtkfgfekindmvpdydrdenrppvkapkkkggkrkrttdeaaeepevd 592

QY 566 --YQOTRUG-----YYPGAKRLTARKRRSRK-----EEVEED 597

Db 593 taadvkdapegpkvpeeeramgpnrvvypigceewlkpkkrevkdkgvpeevde 651

RESULT 5

ID AAY44502

XX AAY44502 standard; Protein; 626 AA.

XX AAY44502;

XX AAY44502;

DT 27-MAR-2000 (first entry)

XX N. crassa

XX delta228-UV damage endonuclease.

XX

KW Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;
KW glutathione-S-transferase signal peptide; uvell+ gene product;
KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
KW abasic site; platinum diaduct; mismatched nucleotide pairing;
KW nucleotide alkylation; skin cancer.

OS Neurospora crassa.

XX WO9963828-A1.

PN 16-DEC-1999.

XX 08-JUN-1999; 99WO-US12910.

XX 08-JUN-1998; 98US-0088521.

PR 18-MAY-1999; 99US-0134752.

XX (UYEM-) UNIV EMORY.

PA Doetsch PW, Kaut B, Avery AM;

XX WPI; 2000-J16417/10.

DR A new truncated ultraviolet damage endonuclease for treatment of skin

XX cancers -

PT Claim 16; Page 59; 133pp; English.

PS The present sequence is delta228-UV damage endonuclease from N. crassa.

XX Delta228-UVDE is a 288 amino acid deletion of the N-terminal of the

CC uvell+ gene product. This is expressed in frame with a GST leader sequence

CC to generate a fusion protein. This provides stable endonuclease fragments

CC for cleaving a double-stranded DNA molecule that has a distorted

CC structure resulting from UV radiation damage, a photoproduct, an abasic

CC site, mismatched nucleotide pairing, platinum diaduct, an intercalated

CC molecule or alkylation of a nucleotide. Uvelp can be used in compositions

CC for internal or topical application and as a therapeutic agent for skin

CC cancers.

XX Sequence 626 AA;

SQ

Query Match 30.9%; Score 961.5; DB 21; Length 626;

Best Local Similarity 36.2%; Pred. No. 5.9e-70;

Matches 233; Conservative 85; Mismatches 145; Indels 181; Gaps 18;

QY 110 ARSSKKKATDSVSDKIDSVASYDSST-----HLRRSSR-----SKPVPVNNSSSESE- 157

Db 3 srkskaaaidtpqsesstfsstldssapsparnlrrsgnllqpsssekdrdhekrsgeel 62

QY 158 -----SSEQISKATKKVKKEEEYVEEVDEK 184

Db 63 agrmmgkdanghclregkeegvkmaleglarmerriqratkrgkqgleed----- 114

QY 185 SLKNSSSDEFPVPVPEOLETPIKRRRS-----RSSAKNLEKESTMNLDHHA----- 232

Db 115 gipvpvsvrft-ptapyhhkstnaeereakepvltkshdkdveraeagvddvkmepaat 173

QY 233 -----PREMFDCLDK-----PIPWGRGLGYACLTILRSKMERVFCSCRTCRIT 275

Db 174 nilepedaqdaaergaarppavpnssylplpwkgrlgyacintylrnsakpplfsrtcma 233

QY 276 TI-----QRDGLSVKQLGTQNVLDLKLVEVNNHFG 307

Db 234 sivdhrhplqfedephehlknkpdkskepqdelghkfvgelglanardivkmlewnesqy 298

QY 308 IHFMVSSDLFFPFASHAKYGYTL-EFAQSHLEEVGKLANKYNHRLTMHPGQYTOIASPRE 366

Db 289 -----fpfashpvhgylapfasevlaeagrvaaelghrltthpgqftqlgsprk 338

QY 367 VVVDASIRDLAYHDLIRSMKLNQNKDAVLIILHGGTFEGKKEKTLDRFRKNYORLSDS 426

Db 339 evvesairdleyhdelisllkpeqndravmihmgqgfdkaatlfrkryarlsqgs 398

Qy 427 VKARVLVLENDVSKSVODLLPLCOELNIPVLVDWHHNIIV--PGLTREGSLDLM--PLIP 482
 Db 399 cknrlvlehdvgtvhdlllpvceelnipmvlldyhhnncfcpahiregtldisdpklqe 458
 Qy 483 TIRETWTRKGITOKOHYSESADPTAISGMKRRRAHSRVRFPDPPCDPTMDLMTAEKKEQA 542
 Db 459 riantwkrkgikgmhysepcd-gavtpdrkrhrprvmtlppcpdmdlmeakdkega 517
 Qy 543 VFELCRRYEL-----QNPPC-----PLEIM 562
 Db 518 vfelmtfklpgfekindmvpdydrddenpappvkapkkkgkgrkrttdeaaapeeve 577
 Qy 563 GPEYDQTRDG-----YYPGAEKRLTARKRRSRK-----EEVEED 597
 Db 578 vpeeeramgpyrnrvywpvgceewlkpkrevkkgvpeevede 621

RESULT 6
 AAY44503
 ID AAY44503 standard; Protein: 626 AA.
 AC AAY44503;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE B. subtilis delta228-UV damage endonuclease.
 KW Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;
 KW glutathione-S-transferase signal peptide; uvcl+ gene product;
 KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
 KW abasic site; platinum diaduct; mismatched nucleotide pairing;
 KW nucleotide alkylation; skin cancer.
 XX
 OS Bacillus subtilis.
 XX
 PN WO9963828-A1.
 XX
 PD 16-DEC-1999.
 XX
 PF 08-JUN-1999; 99WO-US12910.
 XX
 PR 08-JUN-1998; 98US-0088521.
 PR 18-MAY-1999; 99US-0134752.
 XX
 PA (UYEM-) UNIV EMORY.
 XX
 PI Doetsch PW, Kaur B, Avery AM;
 XX
 DR WPI; 2000-116417/10.
 XX
 PT A new truncated ultraviolet damage endonuclease for treatment of skin
 PT cancers -
 PS Claim 16; Page 59; 133pp; English.
 XX
 CC The present sequence is delta228-UV damage endonuclease from B. subtilis.
 CC Delta228-UVDE is a 288 amino acid deletion of the N-terminal of the
 CC uvcl+ gene product. This is expressed in frame with a GST leader sequence
 CC to generate a fusion protein. This provides stable endonuclease fragments
 CC for cleaving a double-stranded DNA molecule that has a distorted
 CC structure resulting from UV radiation damage, a photoproduct, an abasic
 CC site, mismatched nucleotide pairing, platinum diaduct, an intercalated
 CC molecule or alkylation of a nucleotide. Uvclp can be used in compositions
 CC for internal or topical application and as a therapeutic agent for skin
 CC cancers.
 XX
 SQ Sequence 626 AA;

Query Match 30.9%; Score 961.5; DB 21; Length 626;
 Best Local Similarity 36.2%; Pred. No. 5.9e-70;
 Matches 233; Conservative 85; Mismatches 145; Indels 181; Gaps 18;

Qy 110 ARSSKKKATDSVSDKIDESVASYDSST-----HLRRSR-----SKKPVNYSSESE- 157
 Db 3 srkskaaaltdfpqsesstfsstldssapsparnlrgrsnllqpseksdrdrkrsgeel 62
 Qy 158 -----SPEQISKATKKVKQKEEPEYVEVDEK 184
 Db 63 agrmmgkdanghclregkeqegvkmaiegllarmerrlqratrkqkqgleed----- 114
 Qy 185 SUKNSSSEDEFVPEQLETPISKRRS-----RSSAKNLEKESTMLNDHHA----- 232
 Db 115 gippsvsvrftptapyhhkstnaeeraekapvltkshkdvereaeigvddvkmepaat 173
 Qy 233 -----PREMFDCLDK-----PIPWGRILGYACLTWTLILSMKERVPCSTCRIT 275
 Db 174 nliepedaqdaaagaarppavvnssylplpwkgriygaclntylrnakppifssstcrma 233
 Qy 276 TI-----QRDGLVSKOLCTQNVLDLILKLVENHNFG 307
 Db 234 sivdhrhplqfedephehlknkpkdskepqqdelghkfvqelglanardivkmclw----- 288
 Qy 308 IHFMRVSSDLFPFASHAKYGYTL-EFAQSHLEEVGKLANKYNHRLTMHPGOYTQTASPRE 366
 Db 289 -----fpfashpvhgkykiapfasevlaeagrvaaelghritthpgqftqlgsprk 338
 Qy 367 VVDSAIRDLAYHDEILSPMKLINEQLNKDAVLIILHGGTFECKKETLDRFKNYQRLSDS 426
 Db 339 evesaairdleyhdellisliklpeqgnrdavmlihmvgqfkdkaatlerfkrnyarlsgs 398
 Qy 427 VKARVLVLENDVSKSVQDILLPLCQELNIPVLVDWHHNIIV--PGLTREGSLDLM--PLIP 482
 Db 399 cknrlvlehdvgtvhdlllpvceelnipmvlldyhhnncfcpahiregtldisdpklqe 458
 Qy 483 TIRETWTRKGITOKOHYSESADPTAISGMKRRRAHSRVRFPDPPCDPTMDLMTAEKKEQA 542
 Db 459 riantwkrkgikgmhysepcd-gavtpdrkrhrprvmtlppcpdmdlmeakdkega 517
 Qy 543 VFELCRRYEL-----QNPPC-----PLEIM 562
 Db 518 vfelmtfklpgfekindmvpdydrddenpappvkapkkkgkgrkrttdeaaapeeve 577
 Qy 563 GPEYDQTRDG-----YYPGAEKRLTARKRRSRK-----EEVEED 597
 Db 578 vpeeeramgpyrnrvywpvgceewlkpkrevkkgvpeevede 621

RESULT 7
 AAY44505
 ID AAY44505 standard; Protein: 294 AA.
 AC AAY44505;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE D. radiodurans delta228-UV damage endonuclease.
 XX
 KW Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;
 KW glutathione-S-transferase signal peptide; uvcl+ gene product;
 KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
 KW abasic site; platinum diaduct; mismatched nucleotide pairing;
 KW nucleotide alkylation; skin cancer.
 XX
 OS Dienococcus radiodurans.
 XX
 PN WO9963828-A1.
 XX
 PD 16-DEC-1999.
 XX
 PF 08-JUN-1999; 99WO-US12910.
 XX
 PR 08-JUN-1998; 98US-0088521.
 PR 18-MAY-1999; 99US-0134752.
 XX


```

Db 366 -----veekleetygyfrefkfttrilvkrkrnkeqkklkedekkli 408
Qy 509 SG-----MKRRASDRVP-----DFPPCDPTMDLMIKAEKQAVFELCRRYEL 552
Db 409 aaeepddekklkdsdkvvvpvknkssfpd-----kfrapdkrtmf-----yrl 456
Qy 553 QN--PPCP-----LEIMGPEYDQTRGYYPPGAEKRLTA-----RKRRSRKEVEEDE 598
Db 457 selfpiivrkdnelavsgdcmksvng-----kklkstfnfkrnrnklikermqge 507

RESULT 12
AAW54145
ID AAW54145 standard; Protein; 1639 AA.
AC AAW54145;
XX
DT 23-SEP-1998 (first entry)
XX
DE P. falciparum synthetic gp190 protein.
XX
KW gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
KW monoclonal antibody; passive immunisation; parasite.
XX
OS Plasmodium falciparum.
OS Synthetic.
XX
PN WO9814583-A2.
XX
PD 09-APR-1998.
XX
PF 02-OCT-1997; 97WO-EP05441.
XX
PR 02-OCT-1996; 96DE-4040817.
XX
PA (BUJA/) BUJARD H.
XX
PI Bujard H, Pan W, Tolle R;
XX
DR WPI; 1998-240088/21.
DR N-PSDB; AAV21451, AAV35363.
XX
PT Recombinant production of complete gp190/MSP-1 Plasmodium surface
PT protein - useful in anti-malaria vaccines, also stabilising genes by
PT reducing their AT content
XX
PS Example 1; Fig 3c; 48pp; German.
XX
CC This sequence represents a modified Plasmodium falciparum gp190/MSP-1
CC (merozoite surface) protein. The gene encoding this protein has been
CC stabilised by reducing the AT content of the nucleotide sequence. Such a
CC protein is useful in vaccines against malaria or for producing monoclonal
CC antibodies (for passive immunisation). The complete gp190 protein can now
CC be produced outside the parasite and has, at least over extended regions,
CC the native pattern of folding. Larger amounts of the protein can be
CC produced recombinantly than would be possible using the parasites as
CC source.
XX
SQ Sequence 1639 AA;

Query Match 4.4%; Score 136.5; DB 19; Length 1639;
Best Local Similarity 19.5%; Pred. No. 0.066;
Matches 130; Conservative 107; Mismatches 212; Indels 217; Gaps 35;

Qy 17 FTILKQKAFKNHPCVPSVCTITYS---RFHCLPDTLTKSLLPMSKTTLSMLPQVNIGAN 73
Db 185 flllrakl---ndvcandycipfnlknanel-dvllkly----- 221

Qy 74 SFSAPTPVDLKKENETELANTSGPHKSTSTSTRKRARSKKKATDSVSKIDESVASYD 133
Db 222 -fgyrkpldnldkdvngmedyi---kkn-----kktieninelieeskktd 264

```

```

Qy 134 SSTHLRRSSRSKKPVN-----YNSSSESESEIOISKATKKVKOKEEEEYVEEVDK-- 184
Db 265 kknatkkeekkklyqagydsiynkgle-eahmlisvlekrldtlkknenkellidkn 323
Qy 185 SLKN-----ESSDSEFEPVPEQLET-----PISKRRRSRSA 217
Db 324 eiknppansgntpntllidknkkieehekeiaktikfnidsiftpleyleylrokn 383
Qy 218 KULE-----KESYTNLDDHAPREMFDCLDKPIPWGRGUYACLNTPILRSM----- 262
Db 384 knidisakvetkest-----epneyngvtyp-----lsyndnnaelneinsfgdlin 431
Qy 263 -----KERVFCSTCRITTIQDGLSVKQL---GTQNVLDLKLVE-- 301
Db 432 pfdytkepsknlytdnerkfkineikekiesdkksyedrsksinditkeyekl 491
Qy 302 -----WNHFGIHFMRVSSDLFPFAS--HAKYCYTLEFAQSHLEEVGKILANKYNHRL 351
Db 492 lneiydskfnnni-----dltnfekmmgkryskveklthhntfasyenskhlek 542
Qy 352 TMHPCGYTOIASPREVVDSAIRDLAYHDEILSRMK-----LNFOLNKDAVLIHILGGTF 406
Db 543 ltkalkymedyslrnlivve---kelkyyknlskietietlvenikkdeeql-----f 593
Qy 407 EGKKEPTLDRFRKNYQ--RLSDSVKAR---LVLNDDVSVSVQDILLPLCOELNIPILVLDW 460
Db 594 e-kkikdenkpkdekilevsdivkvqkvllmnikidelkkktllik-----nvel--- 643
Qy 461 HHNHI--VPGTLREGS-----LD---LMPLTIPTRETWTTRKGTQKHYSSES 502
Db 644 -knhivpnsyqenkepyyilivikkeldkikvfmkpveslineekkniklegq--sdn 700
Qy 503 ADPTA---ISGMKRRASDRVDFPPCDPTMDLMIKAEKQAVFELCRRYELQNPQPL 559
Db 701 sepstegeitgattkpgqgagsalegd---svqagaqeqkga-----qpvpv 746
Qy 560 EIMGPE 565
Db 747 pv--pe 750

RESULT 13
AAW5657
ID AAY85657 standard; Protein; 1341 AA.
XX
AC AAY85657;
XX
DT 12-FEB-2001 (first entry)
XX
DE Human Acinus L protein sequence.
XX
KW Chromatin aggregation induction; human; Acinus; AIDS; herpes; diabetes;
KW Alzheimer's disease; Parkinson's disease; cardiac infarction;
KW brain infarction; adenovirus infection; viral hepatitis.
XX
OS Homo sapiens.
XX
PN WO200061743-A1.
XX
PD 19-OCT-2000.
XX
PF 07-APR-2000; 2000WO-JP02254.
XX
PR 09-APR-1999; 99JP-0103317.
XX
PA (SHIO ) SHIONOGI & CO LTD.
XX
PI Sahara S, Equichi Y, Tsujimoto Y;
XX
DR WPI; 2000-665130/64.
DR N-PSDB; AAC61196.
XX

```


Db 347 grgtgdpiprakkvppnllkaqegkvrtnsdrrkspgslpkveemdmdddefeqptmsfesy 406
Qy 238 DCLDKPIPWGRGLGYACLNLTILSRMKERVFCSTCRITTIQRDGLSVKOLGT-QNVLDL 296
Db 407 lsydqp-----rkkkkvvktsg---talgekgllkkdkstknlinsa 447
Qy 297 IKLVENHNFGIHF-----RVSSDLFPF-----ASHAKYG--YTFEFAQS-- 335
Db 448 qklpanenksdklpagaeptprkvptdvlpalpdlipaiqtnyrplpslelissfq 507
Qy 336 -----HLEEVGKLANKYNHRLTMHPGQVQTIASPREVVVDSAIRDLAYHDEILSR 385
Db 508 pkrkafspqeeeaagftgrmnmqvygskcaylpkmmthqqcivlknidsidq 567
Qy 386 MKLNEQLN-----KDAVLIHILGGTF-EGKKTLDLRFKKNYORLSDSVKARLVLENDV 438
Db 568 lyrieecnhvlietdqlwkvchrdfkeerpeeyeswremyirlgdareqrlrlitnni 627

RESULT 15

AAW60787
ID AAW60787 standard; protein; 753 AA.

AC AAW60787;

DT 08-SEP-1998 (first entry)

DE Deletion mutant delta546-565 of Elongin A.

KW Elongin A; rat; elongation activity; stimulation; RNA polymerase II;
KW Binding; Elongin B; Elongin C; regulation; transcription activity;
KW RNA polymerase; in vitro transcriptional assay; deletion mutant.

XX Synthetic.

OS Rattus sp.

XX WO9814574-A1.

XX 09-APR-1998.

XX 02-OCT-1997; 97WO-US17992.

XX 04-OCT-1996; 96US-0725459.

XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.

XX Conaway JW, Conaway RC;

XX WPI; 1998-286420/25.

XX Isolated Elongin A and C and fragments - used for regulating
PT transcriptional activation of RNA polymerase, particularly for use
PT in in vitro assays and systems

PS Disclosure; Pages 86-88; 185pp; English.

XX AAW60783-92 represent deletion mutants created to define the Elongin A
CC sequences required for binding to Elongins B and C and for
CC transcriptional activity. Elongin A stimulates the elongation activity
CC of RNA polymerase II. Mutants that span the transcriptional activation
CC domain from residues 400 to 730 (AAW60775) were constructed and
CC expressed in Escherichia coli. The present deletion mutant lacks amino
CC acids 546 to 565 of the wild type Elongin A (AAW60771). The Elongin
CC polypeptides and fragments can be used for regulating the transcriptional
CC activity of RNA polymerase. They can be used in in vitro transcriptional
CC assays or systems.

XX Sequence 753 AA;

Query Match 4.3%; Score 134.5; DB 19; Length 753;
Best Local Similarity 20.3%; Pred. No. 0.03;

Matches 86; Conservative 73; Mismatches 169; Indels 95; Gaps 18;
Qy 73 NSFSAETPVDLKKENETELANISGPHKKSTSTSTKRARSSKKKATDSVSKIDESVASV 132
Db 243 sshkekrrpvdargdekssvmgrekshkasskeesrrll-----sedsakelipsvvkk 296
Qy 133 DSSTHLRRSSRSRKKPVN--YNSSESSESEPEQISKATKKVKQKEEYVEEVEDEKSLKNES 190
Db 297 ekd---reunskklspaldvasdnhfkpkhkdksekikskdnkqsvdsvd-----s 346
Qy 191 SSDEFEPV-----VPEQLETPISKRRRS--RSSAKNLEKESTMNLDDHAPR-----EMF 237
Db 347 grgtgdpiprakkvppnllkaqegkvrtnsdrrkspgslpkveemdmdddefeqptmsfesy 406
Qy 238 DCLDKPIPWGRGLGYACLNLTILSRMKERVFCSTCRITTIQRDGLSVKOLGT-QNVLDL 296
Db 407 lsydqp-----rkkkkvvktsg---talgekgllkkdkstknlinsa 447
Qy 297 IKLVENHNFGIHF-----RVSSDLFPF-----ASHAKYG--YTFEFAQS-- 335
Db 448 qklpanenksdklpagaeptprkvptdvlpalpdlipaiqtnyrplpslelissfq 507
Qy 336 -----HLEEVGKLANKYNHRLTMHPGQ---YTQIA-----SPREVVVDSAIRDL 376
Db 508 pkrkafspqeeeaagftgrmnmqvygskcaylpfvegvgvpyvsvlepvlerctpdq 567
Qy 377 AYHDEILSRMKLNEQLNKDAVLIHILGGTF-EGKKTLDLRFKKNYORLSDSVKARLVLEN 435
Db 568 lyrieecnhvlietdqlwkvchrdfkeerpeeyeswremyirlgdareqrlrlit 624
Qy 436 DDV 438
Db 625 nni 627

Search completed: January 15, 2002, 13:49:59
Job time: 288 sec

